Subrata Trivedi · Hasibur Rehman
Shalini Saggu · Chellasamy Panneerselvam
Sankar K. Ghosh *Editors*

DNA Barcoding and Molecular Phylogeny

Springer

DNA Barcoding and Molecular Phylogeny

Subrata Trivedi • Hasibur Rehman •
Shalini Saggu • Chellasamy Panneerselvam •
Sankar K. Ghosh
Editors

DNA Barcoding and Molecular Phylogeny

Springer

Editors
Subrata Trivedi
Faculty of Science, Department of Biology
University of Tabuk
Tabuk, Saudi Arabia

Hasibur Rehman
Faculty of Science, Department of Biology
University of Tabuk
Tabuk, Saudi Arabia

Departments of Pathology
School of Medicine, The University of Alabama at Birmingham (UAB)
Birmingham, AL, USA

Shalini Saggu
Faculty of Science, Department of Biology
University of Tabuk
Tabuk, Saudi Arabia

Departments of Dermatology
School of Medicine, The University of Alabama at Birmingham (UAB)
Birmingham, AL, USA

Chellasamy Panneerselvam
Faculty of Science, Department of Biology
University of Tabuk
Tabuk, Saudi Arabia

Sankar K. Ghosh
University of Kalyani
Kalyani, West Bengal, India

ISBN 978-3-319-90679-9 ISBN 978-3-319-90680-5 (eBook)
https://doi.org/10.1007/978-3-319-90680-5

Library of Congress Control Number: 2018947792

© Springer International Publishing AG, part of Springer Nature 2018
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Part I DNA Barcoding: Advantages and Significance

DNA Barcoding Significance and Utilities 3
Sambashiva Daravath, Reddya Naik Bannoth, Manickam Tamil Selvi, and Srinivas Ankanagari

R in DNA Barcoding 31
Asim Kumar Mahadani, Pradosh Mahadani, and Goutam Sanyal

Implications and Utility of DNA Barcoding 45
J. Suriya, M. Krishnan, S. Bharathiraja, V. Sekar, and V Sachithanandam

"Significance of DNA Barcoding in Avian Species: Tracing the History and Building the Future" 65
Farhina Pasha

DNA Barcoding: A Potential Tool for Invasive Species Identification . . . 73
Muniyandi Nagarajan, Akash Nambidi Parambath, and Vandana R. Prabhu

Part II DNA Barcoding of Microbes

Microbial DNA Barcoding: Prospects for Discovery and Identification 89
Anand Mohan, Bableen Flora, Madhuri Girdhar, and S. M. Bhatt

DNA Barcoding on Bacteria and Its Application in Infection Management 103
Mohammad Zubair, Farha Fatima, Shamina Begum, and Zahid Hameed Siddiqui

Part III DNA Barcoding in Plants

DNA Barcoding: Implications in Plant-Animal Interactions 123
Muniyandi Nagarajan, Vandana R. Prabhu, Ranganathan Kamalakkannan, and Palatty Allesh Sinu

DNA Barcoding in Forensic Botany 143
Mohamed Rizk Enan

A Molecular Assessment of Red Algae with Reference to the Utility of DNA Barcoding 163
Zahid Hameed Siddiqui, Zahid Khorshid Abbas, Khalid Rehman Hakeem, Mather Ali Khan, and Abdul Ilah

DNA Databases: Promises and Limitations for Plant DNA Barcoding 179
Selvaraj Dhivya, Mohanasundaram Saravanan, and Ramalingam Sathishkumar

Aquatic Plant Biodiversity and DNA Barcoding 197
Sufia Irfan and Aishah Alatawi

Part IV DNA Barcoding in Animals

DNA Barcoding of Mosquito Species 217
Lalita Gupta, Sanjeev Kumar, and Kuldeep Gupta

DNA Barcoding of Rays from the South China Sea 229
B. Akbar John, M. A. Muhamad Asrul, Wahidah Mohd Arshaad, K. C. A. Jalal, and Hassan I. Sheikh

Molecular Phylogeny of Elasmobranchs 245
A. Pavan-Kumar, P. Gireesh-Babu, A. K. Jaiswar, S. G. Raje, A. Chaudhari, and G. Krishna

A Review on DNA Barcoding on Fish Taxonomy in India 259
V. Sachithanandam and P. M. Mohan

Applications of DNA Barcoding in Fisheries 281
A. Pavan-Kumar, A. K. Jaiswar, P. Gireesh-Babu, A. Chaudhari, and G. Krishna

Identification and Conservation of Reptiles Through DNA Barcoding 293
Subrata Trivedi, Hasibur Rehman, Shalini Saggu, Al Thabiani Aziz, Chellasamy Panneerselvam, and Sankar K. Ghosh

DNA Barcoding in Avian Species with Special Reference to Taxonomically Wide Biogeographic Studies 305
Farhina Pasha

Molecular Characterization of Ruminant Mammals Using DNA Barcodes 317
Muniyandi Nagarajan, Koodali Nimisha, and Subhash Thomas

Part V Case Studies

Phylogenetic Diversity of Culturable Marine Actinobacteria Isolated from the Havelock Island, the Andamans, India 333
Gobalakrishnan Rajagopal and Sivakumar Kannan

DNA Barcoding and Molecular Phylogeny of Indigenous Bacteria in Fishes from a Tropical Tidal River in Malaysia 351
Mohammad Mustafizur Rahman, Mohd Haikal Izzuddin, Najmus Sakib Khan, Akbar John, and Mohd Azrul Naim

DNA Barcoding of Ichthyoplankton and Juvenile Fishes of a Tropical River in Malaysia 367
B. Akbar John, Hassan I. Sheikh, K. C. A. Jalal, B. Y. Kamaruzzaman, H. Sanower, M. Nur Hanisah, M. H. Rahman, and M. Rozihan

Genetic Variation of Wild and Hatchery Populations of Climbing Perch, *Anabas testudineus* (Bloch, 1792), in Peninsular Malaysia 383
Ahmad Azfar Mohamed, Siti Waznah Abdurahman, and Akbar John

Molecular Identification of Reptiles from Tabuk Region of Saudi Arabia Through DNA Barcoding: A Case Study 397
Bishal Dhar, Mohua Chakraborty, N. Neelima Devi, Sorokhaibam Malvika, Madhurima Chakraborty, Subrata Trivedi, Abdulhadi A. Aloufi, and Sankar K. Ghosh

***Hippophae rhamnoides* (Sea Buckthorn): A High-Altitude Medicinal and Adaptogenic Plant—Molecular Characterization and Bar-Coding** 413
Shweta Saxena, Om Prakash Chaurasia, and Ratan Kumar

Closing Shots: DNA Barcoding and Molecular Phylogeny 439
Subrata Trivedi, Hasibur Rehman, Shalini Saggu, Chellasamy Panneerselvam, and Sankar K. Ghosh

Part I
DNA Barcoding: Advantages and Significance

DNA Barcoding Significance and Utilities

Sambashiva Daravath, Reddya Naik Bannoth, Manickam Tamil Selvi, and Srinivas Ankanagari

Abstract DNA barcoding is a genetic-based tool and used as an integrated approach with taxonomy for species identification and authentication. In recent years, different genetic and genomic approaches are taken in identifying gene markers for universal applicable DNA barcode in different taxonomic groups and samples. In the postgenomics era, combination of molecular biology techniques, bioinformatics, DNA barcoding, and metabarcoding is giving an opportunity to change the existing use of biodiversity information for basic and practical applications. DNA barcoding projects in different organisms are establishing the reference libraries with known sequences available for open access databases to identify unknown specimens. SOPs in DNA barcoding are important for each species from sampling to analysis by the researchers. DNA barcoding is aiding to catalogue the list of species biodiversity and answer fundamental questions in ecology, evolution, and conservation biology. DNA barcoding can be used by regulatory authorities and have advantages as quality control test. DNA barcoding can also improve quality assurance of industrial products. There is also emerging significance in implementing DNA barcoding in consumer, environmental, health, and agricultural protection.

Keywords DNA barcoding · DNA metabarcoding · Informatics · SOP · Quality control · Biodiversity assessment

S. Daravath
Department of Biotechnology, Nizam College, Hyderabad, Telangana, India

R. Naik Bannoth
Department of Zoology, Osmania University, Hyderabad, Telangana, India

M. T. Selvi
Value Added Corporate Services Pvt Ltd, Chennai, Tamil Nadu, India

S. Ankanagari (✉)
Department of Genetics, Osmania University, Hyderabad, Telangana, India

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_1

1 DNA Barcoding: A Genetic Approach-Based Tool for Species Identification

Developments in molecular genetics and genomic technologies help in easy and reliable identification of species. Researchers are using short DNA sequences, called as DNA barcode, from the specified region of genome. The concept of DNA barcoding was proposed by Hebert et al. (2003a). A DNA barcode is a part of the genome of the organism that stands for standardized target regions of genic and/or intergenic in any biological species. DNA barcoding involves amplifying, sequencing, and querying of information from extranuclear genomes, both chloroplast and mitochondrial. Once the DNA barcoding sequence has been retrieved from an unknown sample, an algorithm is used to compare with unknown sequence to reference database containing known barcodes from identified samples that makes easy to identify the unknown samples. When identification of biological species based on morphological characters is not possible, DNA sequences of highly conserved genes are used. DNA barcoding helps to differentiate similar-looking species and identifying immature or damaged specimens. DNA barcode carries both specific and systematic data. DNA barcode development established an error-free and fast method of species identification and can function as molecular identifier for every species.

1.1 Significance: Biodiversity Assessment

Biodiversity is the variability of life on Earth, including the genes which they carry and complex ecosystems that create the environment (Primack et al. 2001). The degree of biodiversity is constantly changing; the rate of extinction is 100–1000 times higher, almost exclusively as a consequence of human activity. In recent years, the effort to map and to protect global biodiversity has been ever increasing. Thus the global project aimed at the mapping and protection of biodiversity using a new taxonomic method called "DNA barcoding" was started. DNA barcoding assists in cataloging the list of species biodiversity and answers fundamental questions in ecology, evolution, and conservation biology. Many authors have proposed DNA barcoding as an integrated approach with classical taxonomy for species identification and authentication in the postgenomics era (Kane and Cronk 2008; Newmaster et al. 2009; Sahare and Srinivasu 2012; Vohra and Khera 2013). Since it has enough nucleotide diversity, DNA barcodes are used to discover, describe, and understand biodiversity of unknown species. The universal coded labels found by the researchers are used as unique identification marker for the species on the planet (Kress et al. 2015). Combination of molecular biology techniques, bioinformatics, and DNA barcoding is giving an opportunity to change the existing use of biodiversity information for basic and practical applications.

1.2 DNA Barcoding Pipeline

The Barcode of Life (2010) defines four aspects of DNA barcoding workflow. Firstly, verified taxonomic identification and ideally voucher specimens are used as critical sources for DNA library preparation from national parks, botanical gardens, zoological gardens/zoos, seed banks, national herberia, and gene banks. Secondly, these samples are subjected to analysis in the laboratory where extraction of DNA is done as per the laboratory standard operating procedures. After extraction of DNA, amplification is done by using PCR method and sequenced. Thirdly, the obtained gene sequences are kept in the database, a reference library, which helps the allocation of known sequence. Lastly, by using algorithm, best fit reference record is identified from the database for the unknown sample data.

2 Genetic and Genomic Approaches in DNA Barcoding

2.1 DNA Barcoding with Gene Markers

The criteria for identification of gene marker regions for univerisally applicable DNA barcode standard include following: (1) significant species-level genetic variability and divergence, i.e., DNA barcode region should have high interspecific and low intraspecific variability, (2) short sequence length to facilitate DNA extraction and amplification, (3) simple to sequence universal PCR primers, and (4) few easily alignable insertions/deletions.

2.1.1 Gene Markers in Animal DNA Barcoding

In animals, mitochondrial DNA occurs as a single double-helical circular molecule containing 13 protein-coding genes, 2 ribosomal genes, and several tRNAs. Mitochondrial genes are preferred over nuclear genes because mitochondrial genes lack introns; they are generally haploid and exhibit limited recombination. Each cell has several mitochondria, and each mitochondrion contains several complete sets of mitochondrial genes. Thus, when sample tissue is limited, the mitochondrion offers a relatively abundant source of DNA. Mitochondrial *cytochrome c oxidase* subunit 1 (COI) gene was suggested as unique barcode region for animals (Hebert et al. 2003a). COI is a 648 base pair (bp) fragment of the mitochondrial gene cytochrome *c* oxidase 1 (COI; Hebert et al. 2003b). For animals, the mitochondrial cytochrome *c* oxidase I (*COI*) locus appears to satisfy the desired criteria for most groups. Universal primers at the 5′ end of **COI** gene make a 648 bp fragment easy to amplify in a broad spectrum of phyla, and its nucleotide substitution rate allows identifying close and distinct species. This DNA sequence effectively identifies most of the species. Initially, it has been elevated to the status of "the barcode of life" for

identifying animal species (Hebert et al. 2003b). Compared to previous 650 base long segment of the cytochrome c oxidase gene, a 100 base fragment of the original barcode was reported to be effective in identifying archival specimens and potentially useful for all taxa of the eukaryotes DNA barcoding. *COI* region is now widely used for molecular evaluation of diversity, as it has good potential for identifying cryptic species and improving understanding of marine biodiversity. The other mitochondrial genes used as barcode markers include: *cob*, which encodes for apocytochrome b; *cox2* and *cox3*, which encode for the cytochrome oxidase subunits 2 and 3, respectively; and *nad1*, which encodes for NADH dehydrogenase subunit1 and the mitochondrial 16S-rDNA gene.

2.1.2 Gene Markers in Plant DNA Barcoding

Plant DNA barcoding meant for mainly plastid DNA includes maturase K (matK) and ribulose-bisphosphate carboxylase (rbcl) (CBOL Plant Working Group 2009; Burgess et al. 2011). Several plastidial genes, such as the most conserved rpoB, rpoC1, and rbcL or a section of matK, which shows a fast evolution rate, have been proposed as barcode regions (Shaw et al. 2007). Similarly intergenic spacers such as trnH-psbA, atpF-atpH, and psbK-psbI have been proposed as barcode regions (Fazekas et al. 2009). In 2009, the CBoL (Consortium for the Barcode of Life) Plant Working Group (Hollingsworth et al. 2009) suggested the use of two-locus combination of rbcL and matK as core-barcode regions, due to recovery rate of rbcL and the high resolution of matK. For a single primer pair sequencing, matK is difficult to use (Dunning and Savolainen 2010), whereas, despite its limited resolution, rbcL of matk is considered in terms of amplification, sequencing, and alignment and provides a useful backbone in the creation of plant DNA barcode datasets (De Mattia et al. 2012). The trnH-psbA intergenic spacer is direct to amplify and has a high genetic variability among closely related taxa (Bruni et al. 2010; Kress et al. 2015; Shaw et al. 2007). The nuclear ITS region is also considered as a supplementary DNA barcode region (Li et al. 2011). Although there is still debate on the effectiveness of these markers especially when users are dealing with closely related taxa, DNA barcoding showed consistent results when used to identify unknown specimens based on the comparison with reference sequences (Burgess et al. 2011; De Mattia et al. 2012).

The combining of barcodes from multiple loci has been used successfully. Plant DNA barcodes will prove extremely useful for applications like ecological forensics, identification of traded materials, undertaking identifications, and assisting species discovery in some plant groups. The key step is assembling large DNA sample sets representing the Earth's botanical diversity, supported by voucher specimens and indexed *via* DNA sequences which help to use plant DNA for identification of species effectively (Hollingsworth et al. 2011).

2.1.3 Gene Markers in Algal DNA Barcoding

The green alga is an ancient Chlorophyta, which are used as bioindicator for monitoring water and ecosystem. It is used in biotechnological applications such as the production of fuels, chemicals, food, and animal feed. The identification of green microalgae is very difficult by using traditional method in live cultured cells. DNA barcoding method is helpful to identify the 5000 which is not explained earlier. *rbc*L and *nu*ITS1 and *nu*ITS2 markers are used to identify the green microalgae species from neotropical inland waters. Existing available universal primers for ITS1-5.8S-ITS2 region amplification were done for 92% of the samples. But by designing new set of primers for *rbc*L, which sequenced for 96% of the samples, *nu*ITS1 or *nu*ITS2 sequences identified 35% of species using barcode gap calculations. Accuracy of identification is done by *nu*ITS2 compensatory base change (CBC) and ITS1-5.8S-ITS2 region phylogenetic analysis along with morphological inspection. By using *rbc*L sequences only 6% of right species could be identified, however, with the available analytical tools and reference barcodes in the database *nu*ITS1 or *nu*ITS2 can be other useful markers for DNA barcoding of freshwater green microalgae by using availability of analytical tools and reference barcodes for this marker in the database. DNA barcoding pipeline based on *nu*ITS2 is useful for identification of green microalgae species. It is clear that deposition of more taxonomically accurate reference barcodes in curated databases (e.g., BOLD Systems) will help in lessening known tropical species.

2.1.4 Gene Markers in Fungal DNA Barcoding

COI is the default marker adopted by the Consortium for the Barcode of Life for all groups of organisms, including fungi. Most fungi are microscopic and inconspicuous, and many are unculturable. Hence universal primers are required to detect a truly representative profile where COI seems to have many challenges from other candidates like ITS. ITS region is the main barcode region in fungi (Schoch et al. 2012). Internal transcribed spacer (ITS) is used to identify Oomycota, but the default COI marker is helpful to identify closely related species and in *Penicillium* and other fungi. Across the fungal kingdom, ITS was generally considered as higher resolution than LSU in species identification. ITS performance in species discrimination is as close to RNA polymerase II largest subunit (RPB1) as the protein-coding marker. However, ITS region fails to identify the species in certain groups of fungi and thus requires secondary barcode loci for proper identification.

2.1.5 Gene Markers in Protist DNA Barcoding

Protists are a diverse and loose group of disparate eukaryotic microorganisms. They have very simple structural organization, unicellular or multicellular. This simple cellular organization distinguishes the protists from other eukaryotes. They have

evolved before plants, animals, or fungi appeared on Earth. CBOL initiated Protist Working Group and assessed the efforts to identify the barcode regions across all protist lineages, and recently it has a two-step barcoding approach to assess protistan biodiversity. Various protistan DNA barcodes have been proposed; D1–D2 and/or D2–D3 regions at the 5′ end of 28S rDNA have been positively tested in ciliates, saprophytes, and acantharians and are also promising for diatoms. Highly variable V4 subregion on the 18S rRNA gene may serve as a potential candidate for barcoding diatoms (Zimmermann and his group). ITS1 and/or ITS2 rDNA is also commonly utilized in oomycetes, chlorarachniophytes, and green algae. COI also allows morphospecies identification in red and brown algae, dinoflagellates, some raphid diatoms, Euglyphida, naked lobose and shelled amoebae, coccolithophorid haptophytes, and some ciliates. Other group-specific barcodes include the large subunit of the ribulose-1,5-biphosphate carboxylase–oxygenase gene (rbcL) and the chloroplast 23S rRNA gene for photosynthetic protists and spliced leader RNA genes for trypanosomatids. The resolution powers of different protistan DNA barcodes are not compared.

2.1.6 Gene Markers in Microbial DNA Barcoding

Many microbes are difficult to culture in nature. Hence it is eminent to know microbial association with various environments. Gene sequences of 16S rRNA of various environmental samples pave a way to understand a microbial diversity and cataloguing vast diversity of them. Bacterial 16S rRNA gene is considered one of the important markers for soil and marine ammonia oxidizing bacteria. It indicates that 16S rRNA gene is highly conserved for each and every species of bacteria and it can be used as a marker for DNA barcode for different species (Janda and Abbott 2007). COI gene was also used to develop the DNA barcode in 22 species of pathogen (Jones et al. 2013). In *Wolbachia*, a common endosymbiotic bacterium, DNA barcode was developed by using *COI* gene (Smith et al. 2012) and is having overlap areas in eukaryotic DNA barcode. This study confirms that the COI gene can be a DNA barcode marker for bacteria. Chaperonin-60 (cpn60), also known as GroEL and Hsp60, is a molecular chaperone conserved in bacteria. Thus 16S rRNA, COI gene, and cpn60 can be used as markers for developing DNA barcode.

2.2 DNA Metabarcoding with Genomics

2.2.1 Metagenomics and NGS Technologies

Metagenomics is the study of organisms in a community based on analyzing the DNA within an environmental sample. Examples are profiling of microbes from deep ocean water and soil from mine. These samples data are used for agricultural microbiome analysis, ecological remediation, or other biological investigations.

Environmental metagenomics as an area is very limited prior to the advent of next-generation sequencing (NGS). NGS provided researchers the capability to profile entire microbial communities from complex samples, discover new organisms, and explore the dynamic nature of microbial populations under changing conditions. Metagenomics study also helps us to understand the diversity and possible physiological association of plants with endophytes/microbial communities.

NGS technologies can generate several hundred thousands of millions of sequencing reads in parallel. This can be generated from fragmented libraries of a specific genome (i.e., genome sequencing) or from a pool of PCR-amplified molecules. NGS technologies are helpful in achieving mass parallel sequencing of single DNA molecules, resulting in high-throughput data. Parallel sequencing means all regions and variants of the gene are sequenced to detect pseudo genes, contaminants, or allelic variation (Shokralla et al. 2014). Different methods are used to generate multilocus sequence data for various applications such as genotyping and phylogenetics. One method is amplicon sequencing (McCormack et al. 2013). This multiplexing capability makes NGS a cost-effective method for DNA barcoding. Researchers used both traditional Sanger sequencing and NGS amplicon sequencing to characterize a region of *ITS2* known to contain microsatellites and indels in mosquitoes. The success is compared in both methods in order to determine their applicability in sequencing hypervariable genes such as *ITS2*. The variability found within *ITS2* is also analyzed, and its utility as a DNA barcoding marker for mosquitoes is assessed, compared with previous estimates based on the mitochondrial DNA locus *COI* (Batovska et al. 2016).

2.2.2 DNA Metabarcoding

Metabarcoding approach relies on parallel throughput sequencing of reads from fragmented libraries of a specific genome (i.e., genome sequencing) or from a pool of PCR-amplified molecules (i.e., amplicon sequencing (Pavan-Kumar et al. 2015)). Metabarcoding combines two technologies: DNA-based identification and high-throughput DNA sequencing. It uses universal PCR primers to mass amplify DNA barcodes from mass collections of organisms or from environmental DNA. The metabarcoding can be used in genes (used as barcodes) that are sequenced without a necessity for cloning. Metabarcode datasets are taxonomically more comprehensive, many times quicker to produce, and less reliant on taxonomic expertise. This advanced high-throughput NGS technology facilitates potentially the generation of DNA barcodes faster at a low cost (Shokralla et al. 2014). This is mainly used to obtain whole-genome DNA sequence information from mass environmental samples.

DNA metabarcoding helps to identify the species from the DNA samples by sequencing nucleotide barcodes. Barcodes are detected by monitoring alignment of sequences and identify a pair of conserved regions which is close to the variable one. Commonly used markers are 16S for bacteria, mt16S for mammals, CO1 for insects, ITS1 for fungi, and rbcL, trnl, and matK for plants, although non-conventional

markers can also be used. The data generated using metagenomics provides additional genomic information to identify the taxa and functional characterization of the environment.

Sequencing methodologies for DNA barcoding have changed over time. More recent HTS-based studies have employed one of three major sequencing library preparation strategies: ligation-based "tagmentation" kits (Richardson et al. 2015a, b), singly indexed barcoded primers (Valentini et al. 2010; Hawkins et al. 2015; Keller et al. 2015; Kraaijeveld et al. 2015), or dual-indexed barcoded primers (Sickel et al. 2015). The dual-indexing approach described in Sickel et al. (2015), adapted from Kozich et al. (2013), shows great promise for facilitating library preparation for large studies while reducing multiplexing cost and increasing laboratory efficiency. Sickel et al. (2015) report 2000–3000 high-quality reads to be adequate to describe bee-collected samples with up to 80 taxa. Lastly, pollen metabarcoding has been successfully conducted using a variety of platforms including Ion Torrent (Kraaijeveld et al. 2015), Roche 454 (Valentini et al. 2010; Hawkins et al. 2015; Keller et al. 2015), and Illumina (Richardson et al. 2015a, b; Sickel et al. 2015); however, with increased relative throughput, accurate homopolymer sequencing, and increasing length capabilities, Illumina is the current platform of choice for PCR-based approaches.

NGS is the most standard tool for characterizing the ITS2 (DNA marker for mosquitoes) region in mosquitoes; this can be used in many other insect species and genera. Multiplexing made NGS more efficient in sequencing polymorphic regions successfully and in understanding large diversity of ITS2 alleles present in mosquitoes. DNA barcoding marker ITS2 was able to separate all of the species, apart from members of the *Culex pipiens* complex, given similar resolution as cytochrome oxidase I (COI). The best DNA marker could successfully separate all species and provide a good linkage among species phylogenies. The gene with best resolution can be used for bulk DNA barcoding, where large pools of mosquitoes could be sequenced and identified (Hajibabaei et al. 2011).

Microbes in marine environments are studied by using NGS technology. Analyses of bacterial studies were done by using 18S rDNA (Huber et al. 2007) and 16S rDNA (Sogin et al. 2006). The amplicons, microbe's gene expression in surface seawater (Frias-Lopez et al. 2008), transcriptomic sequencing analysis of cDNA libraries, and functional assemblages within seawater were investigated in bacterioplankton (Mou et al. 2008) were investigated. Marine eukaryotic microbiota was studied through NGS analysis of 18S rDNA amplicons (Stoeck et al. 2010). A shotgun sequencing approach was employed to investigate microbial diversity in seawater (Williamson et al. 2008).

Metabarcoding is a tool to assess biodiversity. Very recently metabarcoding is validated by testing it against three high-quality standard datasets from Malaysia (tropical), China (subtropical), and the United Kingdom (temperate) and that has 55,813 arthropod and bird specimens identified to species level. It can be applied in restoration ecology and systematic conservation planning, minimizing false-positive assignments (Matsen et al. 2010; Zhang et al. 2012) linked with the end users of biodiversity data (Cook et al. 2013). In recent years, using DNA metabarcoding with

NGS of DNA barcode is helping to identify simultaneously multiple species in complex samples.

Only 5% of microbial community was studied among 1.5 million of microbes and fungi. Understanding of microbial diversity at its different habitats was addressed effectively by using metagenomics and metabarcoding. Metabarcoding helps to understand complex microbial communities by single-cell sequencing. Metabarcoding can also be used in the assessment of diet, population genetics, and gut parasites in fecal samples. Metabarcoding offers better taxonomic resolution of food plant species by using multiple genetic markers.

Metabarcoding helps us to monitor biodiversity and is also used to study the patterns of arthropod litter diversity and composition in the tropics. The utility of DNA metabarcoding is applied to study the patterns of litter arthropod diversity across land-use types in Xishuangbanna, China, and the study findings have shown that MiSeq platform was effective for arthropod taxa identification in tropics using 400 bp fragment of the COI gene. To enhance the utility of metabarcoding for large-scale and long-term biodiversity monitoring, it is important to increase the identification and standard barcoding of species, especially in the highly diverse tropics.

Metabarcoding approach makes it possible to identify specific microbes and combinations of microbes which is important for the quality of a given wine in a given vineyard over time. The consistency of the microbial diversity over time may ultimately contribute to the quality of a wine and the reputation of vineyards. Certain combination of microbes over a period of time decides the wine quality it is possible to identify through metabarcoding approach. The consistency of microbial species between regions also contributes the taste differences in wine quality.

2.2.3 Extended DNA Barcode

DNA extracts have a mixture of plastid and mitochondria and generate the sequence data across all the mentioned organelle. At low sequence coverage, approx. 1 GB of data, the genome can be "skimmed" and permits near completion assembly of the high-copy plastid, mitochondria, and ribosomal RNA. Skimming approach has the capacity to make highly fragmented nuclear genome assembly. Genome skimming also helps to prepare single insert-size library for less number of samples, e.g., Illumina Nextera and TruSeq. Larger-size sample library preparation can be automated, e.g., Illumina NeoPrep. It also reduces the cost of analysis. Genome skimming generates some low coverage data of single-copy nuclear regions and can be used in combination with algorithms inspired by Maillet et al. (2014) or Fan et al. (2015) to generate similarity indices between pairs of nuclear genomes which can solve problems in hybridization and/or recent origins. Genome skimming has great promise for extending the plant barcode, which can be used in highly degraded DNAs from herbarium specimens. Sequence data is recoverable from herbarium specimens which has degraded DNA in the absence of PCR, e.g., whole plastid genome from 100-year-old herbarium (Besnard et al. 2014). This absence of PCR is

used when universal primers are ineffective (e.g., *matK* from various plant lineages or plastid regions from many parasitic plants).

2.3 Alternative Next-Generation DNA Barcodes

Several genome simplification techniques have the potential to be implemented as alternative species identification tools. RAD sequencing (Baird et al. 2008) is a commonly used method in population genetics to tackle the limitations of the standard DNA barcoding method and is used to distinguish closely related species (Hohenlohe et al. 2011). It is very much effective at generating sequence data from many thousands of nuclear loci, whereas it is not taxon-specific.

3 DNA Barcoding Informatics

3.1 Reference Libraries

Several studies carried out to date have highlighted many major challenges, including lack of reference libraries, unavailability of the vouchers to professionally identified specimens archived in a herbarium corresponding to the reference DNA sequences in the GenBank (consequently a GenBank reference sequence may be from an incorrectly identified plant species with no way to verify its specific origin), and variable rate of evolution corresponding to different loci (Newsmaster et al. 2013). Researchers have to establish the DNA barcoding library data base with known sequences available for open access to identify unknown specimen.

Development of biological reference material (BRM) acts as a universal platform for reference sequence database at species level to identify plant components in the herbal products. This would consist of taxonomically validated herbarium vouchers of known provenance. The barcode of an unknown sample from commercially produced herbal products can be compared with the reference barcodes to identify the related species. The BRM herbal barcode library presents a method for good manufacturing practices (GMP) of herbal products.

3.2 DNA Barcode Databases

Two central DNA databases are available; one is BOLD (Barcode of Life Data Systems), through which researchers can upload and use it for their analysis and assembling data. BOLD is connected to other databases of voucher specimens by BARCODE Data Standard. Another one are GenBank, EMBL, and DDBJ which are

public repositories of DNA barcode databases maintained by International Nucleotide Sequence Database Collaboration.

3.2.1 BOLD

The Barcode of Life Data Systems (BOLD) is a web-based workbench and database supporting the acquisition, storage, analysis, and publication of DNA barcode records. By assembling molecular, morphological, and distributional data, it bridges a traditional bioinformatics chasm.

BOLD contains now more than 4.2 million validated barcodes (http://www.boldsystems.org/index.php/databases; Ratnasingham and Hebert 2007).

3.2.2 NCBI

National Center for Biotechnology Information (NCBI), a branch of the National Institute of Health in the United States National Library of Medicine (NLM), houses a series of databases relevant to biotechnology and biomedicine and an important resource for bioinformatics tools and services. Major database includes GenBank for DNA sequences. NCBI also collaborates with EMBL and DDBJ which together comprise the International Nucleotide Sequence Database Collaboration and are permanent public repositories for barcode data records. Unknown samples are assigned to a known species by finding the closest database sequence to the sample sequence with nearest neighbor algorithms. The common matching tool available at NCBI, Basic Local Alignment Search Tool (BLAST), employs this algorithm and searches for correspondence between a query sequence and a library sequence.

3.3 DNA Barcoding Projects

Globally several DNA barcoding projects are launched to aim specific taxonomic groups such as plants, fungi, protists, bacteria, and different entities of the animal kingdom including fishes, brides, insects, nematodes, mammals, etc. The largest consortia in this DNA barcoding projects are:

iBOL
A milestone in the field of DNA barcoding was achieved by launching of International Barcode of Life (iBOL) project. The iBOL is the largest biodiversity genomics program. From 25 nations biodiversity scientists, genomics specialists, technologists, and ethicists are working together to construct a richly parameterized DNA barcode reference library which is going to be the basis for a DNA-based identification system for all multicellular organisms. In the initial phase of the operations, iBOL collaborators are barcoding for 5 million specimens representing 500,000

species. During construction of the barcode library, iBOL participants will also be building the infrastructure needed to use it in real-world situations such as conservation, ecosystem monitoring, forensics, and control of agricultural pests and invasive species.

CBOL

To support worldwide DNA barcoding and international online data management system—Consortium for the Barcode of Life (CBOL) was established. Later Barcode of Life Data Systems (http://www.barcodinglife.org) came into effect. Canada was the first country to establish national network for DNA barcoding in Canada initially (BOLNET.ca). Later most of the countries and regions have also established barcoding networks as part of the iBOL, e.g., Europe (ECBOL; http://www.ecbol.org/), Norway (NorBOL; http://dnabarcoding.no/en/), Mexico (MexBOL; http://www.mexbol.org/), and Japan (JBOLI; http://www.jboli.org/).

ECBOL

Earlier times ECBOL was within EDIT, the European Distributed Institute of Taxonomy, to complement CBOL's activities from a European perspective. The idea of ECBOL.org was created within EDIT work package 3.4 "DNA barcoding" and is an information and coordination hub maintained by the Centraalbureau voor Schimmelcultures (CBS) in Utrecht. As a result of EDIT and the DNA Barcoding in Europe meeting in 2007, the ECBOL initiative was started, for a network of European leading labs in the field of DNA barcoding.

Currently, as many as 1,371,809 DNA barcodes are identified and stored, which correspond to approximately 113,435 denominated species (as of September 30, 2011). These projects enhance the information resources available in the genome databases of GenBank, EMBL, DDBJ, and encyclopedia. The ultimate concept of these entire project missions is to develop the DNA barcoding ease and affordable molecular tool for taxonomic research but also to study, preserve, and protect the biodiversity. The applications of this tool in turn benefits science and society.

4 Standard Operating Procedures and Quality Assurance

4.1 Standard Operating Procedures

A standard operating procedure, or SOP, is a set of step-by-step instructions compiled to help personnel to carry out routine operations. SOPs aim to achieve efficiency, quality output, and uniformity of the process or activity performance; it reduces any mislead and failure to comply with the given process. SOP refers to unique procedures, which are not necessarily standard to another species especially in DNA barcoding method. "Standard" could imply that there is one procedure to be followed for that particular species. The focus is always set on repetition of unchanged processes, procedures, and its documentation. The quality assurance unit is responsible for monitoring the compliance of set norms in the performed

activities including test reports and SOPs. Procedures are extensively employed to assist with working safely. They are sometimes called safe work methods statements (SWMS). They are usually preceded by various methods of analyzing tasks or jobs to be performed in a workplace, including an approach called job safety analysis, in which hazards are identified and their control methods described. Procedures must be suited to the literacy levels of the user, and as part of this, the readability of procedures is important. Thus, SOPs ensures the quality of uniformity in working process with a rich outcome of the activity.

The procedures involved in DNA barcoding are sample collection, isolation of DNA, amplification of the barcoding gene, and finally sequencing and analyzing the results. Product obtained by PCR is sent for DNA sequencing and analyzed using both the DNA BOLD and NCBI databases (Jacque Keele et al. 2014). It helps to identify species from samples of DNA from fish, birds, mammals, plants, and invertebrates. Due to poor DNA or failure of PCR, result might go wrong, but PCR with RDLES is providing researchers a way to confirm through molecular biology the identification of organisms.

To enhance the quality of these methods, several laboratories and organizations around the world are getting ISO 17025 accreditation for the methods of DNA sequencing, next-generation sequencing, and PCR ISO 15189. Main focus of ISO 17025 is on the accredited test and calibration method including (1) detailed sampling procedures, (2) specific handling of test and calibration equipment, (3) ISO 17025-compliant result reports and result reporting, and (4) the participation in proficiency testing programs/lab comparison tests. The compliance with ISO 17025 is secured through accreditation by a national accreditation body and independent internal audits. Quality assurance perspectives are crucial. It is vital to that working with small amounts of DNA is challenging, DNA is not visible to the naked eye throughout the process, from sampling to analysis. Also, obtained results in one study cannot be translated directly to other studies that use different primers, sampling methods, lab and field protocols, etc. Thus the entire SOP is important for each species from sampling to analysis, and also each SOP undergoes method verification before implementation in to other projects.

4.2 DNA Barcoding as Quality Control Test

Ecotoxicology laboratories need to guarantee the homogeneity of cultures and the species identification of test specimens through the accuracy and comparability of results. It can be improved by using DNA barcoding quality measure. International standardization organizations (e.g., ISO, OECD) are actively involved to standardize DNA barcoding as a quality control test in ecotoxicology guidelines to include genetic characterization of invertebrate and plant species in terrestrial ecotoxicological tests.

4.3 Quality Assurance

4.3.1 Plant Samples

In plants, successful PCR of barcoding regions is often inhibited by the presence of secondary metabolites. By changing the extraction methods, primer sequences and the use of an engineered polymerase can usually solve the issues. Another area of concern is the use of only plastid barcode regions which have insufficient nucleotide sequence variability to distinguish closely related species. Combining different barcode regions has proven successful in certain cases, but identity of closely related complex groups is still uncertain. Another major problem is mixture of multiple species in the herbal product due to varied PCR success of the selected gene in samples with potentially degraded DNA due to varied gene copy number and PCR bias (Fazekas et al. 2009). Further improvement/development, particularly designing of novel universal primers with the development of BRM herbal barcode library, is required for barcoding vast plant species.

Due to certain limiting factors such as low PCR efficiency, gene deletion, and inadequate variation, no single-locus barcode exists as a universal DNA barcode for plants. To identify the species successfully, multilocus approach can be adopted for barcoding land plants. By combining the universality, discriminatory power, and amplification success of each locus, high discrimination-oriented results can be obtained. The newer trends being followed in plant DNA barcoding with the approaches of NGS and HRM are proving to be immensely helpful in authenticating the useful medicinal plants for herbal drug preparations. These analyses are widely being used in many researches for detection of contamination in herbal mixtures. It needs to be fully validated to identify plant species in herbal medicine.

Plants are influenced by their genome and their environment. Plant metabolism (mainly secondary metabolism, which is mainly responsible for the medicinal properties) is dependent on its environment (Briskin 2000). Thus, merely relying only on the genome-based authentication will be insufficient for quality control of herbal products. Characterization for morphological and biochemical traits will also continue to play its parallel role in identification and assessment of medicinal plants in herbal industry (DeSalle 2006). Thus, DNA barcoding needs to be improved with suitable molecular biology and analytical chemistry tools, in case it is used for species identity. By including systems biology components encompassing genomics (DNA barcoding) and metabolomics (for active secondary metabolites) in a big way and supplemented with need-based use of transcriptomics [specific expressio subset analysis (SESA)] and proteomics (specific proteome) can improve DNA barcoding in plant materials and their usage in herbal medicine.

5 DNA Barcoding Utilities

5.1 Basic Utilities

5.1.1 Plant Biodiversity Assessment

Correct species identification is necessary in many Sapotaceae species. Getting intact floral samples is not always possible. In that context DNA barcoding is helpful for accurate identification of species in several groups of plants. Along with rbcL, matK, and trnH-psbA markers in plants, additional markers such as ITS, nuclear ribosomal transcribed spacer is used to identify Sapotaceae species (Vivas et al. 2014).

Plant–Pollinator Interactions Over Space and Time The movement of pollen is of importance to the long-term structure and function of plant communities, whether natural or managed (Ricketts et al. 2008; Jordano 2010). Advances in pollen DNA metabarcoding afford researchers a more highly resolved understanding of pollination biology at broader scales and greater sampling intensities than previously possible, by enabling characterization of complex pollen assemblages collected from either pollinators or plant stigmatic surfaces. Further, pollen metabarcoding enhances taxonomic resolution and sensitivity for studies investigating ecological phenomena such as plant–pollinator networks, facilitation or competition between plants and pollinators, and plant biogeography.

Additionally, broad patterns in plant–pollinator biology over space and time could be explored. These include application to anthropogenic environmental landscape changes, such as habitat fragmentation (Brosi et al. 2007) and climate change (Inouye 2008; Hegland et al. 2009). Such work could particularly take advantage of historical specimen collections, especially of insect pollinators, many of which were carrying pollen when collected. Specimens are typically labeled with descriptors such as date and place of collection, the collector, and sometimes their association, such as the plants on which they were collected (Pennisi 2000). The use of DNA metabarcoding to track changes in plant communities and plant–pollinator interactions, over time, is valuable in terms of its potential to generate results with conservation implications, such as community reference states for ecological restoration projects.

Ancient Pollen DNA Barcoding Ancient DNA method is used to identify pollen accurately over traditional microscopy-based methods. Most ancient DNA barcoding studies have used sedimentary ancient DNA (*sed*aDNA) to provide complementary data to macrofossil identification and classic palynology (Jorgensen et al. 2012a; Parducci et al. 2013; Pedersen et al. 2013). Parducci et al. (2013) showed that *sed*aDNA metabarcoding method is identifying plants that produce restricted amounts of pollen, taxa that are difficult to identify with palynology. Within pollen grains, DNA can be preserved for millennia if environmental conditions are suitable. Using specific dyes in fossil pollen confirmed the presence of DNA and DNA was extracted from 150,000-year-old pollen (Suyama et al. 1996).

Standard DNA barcoding methods are used by paleoecologists; they use "mini-barcodes" with shorter amplicons (Jorgensen et al. 2012a, b; Parducci et al. 2013). P6 loop of the *trn*L intron in the chloroplast mini-barcode is most commonly used (Taberlet et al. 2007). Using multiple markers provides detailed picture of past vegetation. For example, the *trn*L mini-barcode provides accurate resolution at family level (Taberlet et al. 2007). Similar levels of resolution have been noted for mini-barcodes based on *rbcL* (Little 2014). Plant mini-barcode marker development is a key in terms of increasing the accuracy of taxon identification in ancient DNA.

Melissopalynology Amplicon sequencing on DNA was done in isolated samples of pollen collected by honeybees. To obtain taxonomic identification at the genus level, the ITS2 region is suitable to determine the pollen taxa collected by honeybee colonies located in an area of corn and soybean cultivation, a context in which nutritional stress due to agricultural development would be plausible. Comparison was done with results of metabarcoding approach and traditional microscopic analysis of pollen samples. The internal transcribed spacer (ITS) is used as a genetic barcode to identify monospecific pollen collected by a specialist bee (Wilson et al. 2010). Applied rbcL and trnH-psbA amplicon cloning to characterize the taxonomic composition of bee-collected pollen samples are from Italian Alpine habitats, e.g., *Lonicera* and Lamiaceae (Galimberti et al. 2014). The metabarcoding approach helps in the easy identification of pollen when compared to microscopy. This method is also applied for floral surveying and pollinator habitat preservation. Melissopalynology has been considered difficult previously, but the development of molecular melissopalynological techniques (DNA metabarcoding) has accomplished to overcoming these difficulties.

DNA barcoding is considered more secure compared to traditional method melissopalynology (Bruni et al. 2015). To identify different plant species found in Corsican honey, real-time PCR is used (Laube et al. 2010; Galimberti et al. 2014; Bruni et al. 2015). Amplified plastid markers of *rbc*L and *trnH-psbA* were used to identify the floral composition of honey from the Italian Alps. Trial of pyrosequencing amplicons of the *trnL* (UAA) intron was done to characterize two commercial honeys (Valentini et al. 2010). Hawkins et al. (2015) used the *rbcL* marker and 454 pyrosequencing to characterize 9 honeys. It reveals that DNA metabarcoding is provided much greater levels in terms of identifying species.

Floristic Studies DNA barcoding assists to verify the origin of timber to prevent illegal timber trading busing the standard *rbcL+matK* barcode in *Dalbergia* species. In Indochina, *Dalbergia* is considered valuable timber. DNA barcoding is assisting to verify the origin of timber to prevent illegal timber trading using the standard *rbcL+matK* barcode in this species. DNA barcoding is also used to identify specimens and species distributions in floristic studies.

Cryptic Species DNA barcoding is a possible taxonomic tool and helpful in identifying species fast and accurate in plants. It also helps to find out new cryptic species.

5.1.2 Animal Biodiversity Assessment

DNA barcoding is a critical tool that helps to identify an animal species and gives its genetic sequence database. On comparison of nucleotide distance between species of different genera, it was found that *Sarotherodon melanotheron* and *Coptodon zilli* are closely related. Using cytochrome oxidase subunit I (COI) confirmed accurate identification of these fish species from Southwest Nigeria (Falade et al. 2016).

DNA barcoding is advance in accurate species identification and also allows for the recognition and documentation of unknown taxa across alpha and beta diversity estimation in the tropical bat. DNA barcoding surveys have shown that a more number of species-level taxa are unnoticed by traditional methods. Bat tissue DNA was extracted and the COI barcode region amplified and sequenced and identified nine species-level taxa within samples, based on analysis of the DNA barcodes. The study outcome has shown that high diversity of bats within Peninsular Malaysia (9 species in 13 samples) was identified and demonstrated how DNA barcoding helps and allows for cataloguing and documenting known taxa lacking formal taxonomic status (Wilson et al. 2014).

Evolution Parallel evolution of species were identified at inter and intraspecific level by using three aphid and two buchnera genes eg. *Mollithrichosiphum* and *Buchnera* at intraspecific as well as the interspecific levels. It supports using endosymbiont genes to study host evolutionary history and biogeographical patterns. In addition, Buchnera *gnd* gene acts as a barcoding marker for aphid identification (Liu et al. 2013). The study indicates that the Buchnera gnd gene is as good as COI as a barcoding marker for aphids (Lebonah et al. 2014).

Ecology Rapid habitat loss and degradation are responsible for population decline in a growing number of species. Understanding the natural history of the species is important for designing conservation strategies to enhance habitat or ex situ conservation. Globally ecosystems have undergone rapid changes during the recent past, especially anthropogenic climate change, biodiversity loss, and biological invasions. The immense environmental degradation stresses the need for quick technique for quantifying and monitoring the spatial and temporal dynamics of biodiversity.

In ecological studies DNA barcoding is used to survey the diets of animals by analyzing the fecal matter to identify the remains of the plants (Valentini et al. 2009). Many studies have used NGS technology in diet analysis and in the investigation of gut microbial ecology. Several studies have been conducted on the effect of diet on the gut microbiome of mice using 16S rDNA amplicons. Recently, diet of bats was studied using short COI amplicons to identify the species which allows or permits to understand the relationship of the diet of sympatric cryptic species (Razgour et al. 2011).

Conservation Unknown species identification helps industries of mining, fisheries, and forestry to conserve and manage its environment which in turn gives us economic benefits. To identify all species in the planet will take 2000 years by using traditional taxonomy. By the help of DNA barcoding, it can increase the

identification rate, and in turn it gives scientific and economic benefits. Genetic marker system in barcoding will help to identify accurately and to track endangered valuable species. Applications of DNA barcoding in extinct species can identify hunted wildlife for the traditional medicine, collections of rare species for private parties, and harvest for other products of wildlife.

Biodiversity Genomic studies in conjugation with DNA barcoding can be very effective in assessment of global biodiversity. Canada has a global hub for DNA barcoding. Advanced technology in high-throughput sequencing is used in DNA barcoding of large environmental samples to identify the species. DNA barcoding was effective in analyzing zooplankton sample collected from Equatorial Pacific Ocean (Machida et al. 2009; Ficetola et al. 2008). NGS approach with conventional Sanger sequencing of cytochrome *b* amplicons is used to identify the presence of bullfrogs in freshwater samples.

5.1.3 Microbial Biodiversity Assessment

Microbial samples collected from the burned and unburned land soil sites and the gene used for DNA barcoding analysis were the 16S rRNA gene (Natalie 2013). Microbial samples collected from the burned and unburned land soil sites were analysed for DNA barcoding with 16S rRNA gene. The data has shown that the microbes found in the unburned samples are less prominent indicating that many of soil Archaea microbes have moved quickly after the fire and stablized (Natalie 2013). Barcoding of selected enterobacterial species using fluorescent proteins provides a simple and speedy method for distinguishing species identity (Thao et al. 2013).

In oceans, microbial life is accountable for 98% of the primary production (Sogin et al. 2006). Microbes cause numerous diseases (Galimberti et al. 2013). The increasing community genomics studies and the metagenomics approaches assure real insights on prokaryote biodiversity and molecular evolution. Microbes relied heavily on gene-centric metagenomic profiling using two genes (16S rRNA and 60 kDa chaperonin protein (cpn60)) to recognize and identify the bacteria. Links et al. (2012) evaluated DNA barcodes for bacteria from the 16S rRNA gene and the protein coding cpn60 gene. Assembling consensus sequences for barcodes was shown to be a reliable method for the tracking and identification of novel microbes in metagenomic studies. The most commonly used barcode gene of bacterial and archaeal community is 16S rRNA marker as a barcode to quantify microbial community structure from environmental samples based on DNA sequences. To examine protist diversity in freshwater samples, amplicons of 18S rDNA were used (Medinger et al. 2010).

5.2 *Applied Utilities*

5.2.1 Environment Protection

Sustaining Natural Resources Natural resource managers can monitor illegal trade of natural products made of hardwood trees. FISH-BOL is reference barcode library for hardwood trees to improve management and conservation of natural resources. Protecting endangered species of primate population is reduced in Africa by 90% because of bushmeat hunting. Law enforcement is using DNA barcoding to identify the species of bushmeat in local markets in order to protect the primate populations.

Water Quality Quality of drinking water is monitored by studying organisms health living in lakes, rivers and streams, their health can be measured. DNA barcoding is used to create a library of these species which is difficult to identify. Barcoding can be used by environmental agencies to improve determination of quality and to create better policies which can ensure safe supply of drinking water.

5.2.2 Health Protection

Airborne Allergen Monitoring Plant pollen is one of the major allergens which affects the economy of individual person and work (D'Amato et al. 2005; Davies et al. 2015). Symptoms from this pollen vary based on its origin and its concentration. Pollen monitoring programs are available which are volumetric pollen samplers, whirling arm samplers, or passive samplers. Hirst-type volumetric samplers are used frequently by national pollen monitoring networks that immobilize air particulates on sticky tape (Scheifinger et al. 2013). DNA metabarcoding has been used to identify species of airborne pollen successfully for the monitoring of allergens in the Netherlands (Kraaijeveld et al. 2015). Pollen samples were obtained by volumetric samplers, and an increased taxonomic resolution among classifications was achieved using DNA metabarcoding, while microscopy could only identify pollen to family level in many cases. In addition, aerobiological monitoring not only identifies pollen but also identifies its taxa for occurrences of certain bacteria or fungi of health interest.

Epidemiology Epidemiologists can use DNA barcoding to zero the effects of mosquitoes, tsetse flies, or the feces of migratory birds most responsible for the spread of infectious disease. This helps not only epidemiologists to take more effective action to protect human health; it can also help limit negative side effects of activities of spraying of pesticides. And it gives confidence to tropical countries combating endemic diseases and gives caution to outbreak of possible epidemics (Muturi et al. 2011). To identify disease vectors, DNA barcoding allows nonecologists to identify the vector species that can cause serious infectious diseases to animals and humans and helps to understand the mentioned diseases and its cure. A global initiative of building

a mosquito barcoding reference library helps public health officials to control vector-borne diseases more effectively and with very less use of insecticides.

Drug Resistance DNA barcoding helps to trace beneficial mutations in 25,000 yeast lineages in a larger community. Through NGS, scientists were able to know which lineage and mutations were increased and decrease at 8th generation of yeasts. Sherlock and his colleagues calculated that about 25,000 beneficial mutations with a fitness effect of more than 2% were established by generation 112. The conclusion is that with the time period taken for beneficial mutation, the population of yeast is established and the fitness is due to the influence of mutation. A high-resolution lineage-tracking tool can be used to track pathogenic microbes, cancer cell lines, and animal tumor models. By combining whole genome sequencing, lineage and tracking a powerful method is offered to characterize the mutational spectrum underlying evolution, disease progression, and drug resistance (Genome Web).

5.2.3 Consumer Protection

Food Provenance and Quality Pollen DNA metabarcoding has application in the studies of food provenance and quality. Pollen is a nearly omnipresent environmental biomarker, and most foodstuffs are likely to contain pollen that can be used to trace the product's potential time of origin and its geography. This application is used to trace the geographic and botanical origins of honey from flowers. Honey is a high-value nutritional product and its taste, food quality, and safety differ depending on the plants the honeybees have foraged upon (Crane 1975).

Product labeling guidelines are necessary to prevent mislabeling and often require the floral source of commercially sold honey to be declared (Bruni et al. 2015). Honeys are labeled as monofloral or multifloral by nectar and pollen from a single or multi-plant species. Honeys are classified as monofloral if the pollen content of one species is greater than 45% (Anklam 1998). Honey from one plant species often have higher commerical value and therefore have a greater possibility of adulterations and incorrect labeling (Persano Oddo and Bogdanov 2004). Safety and quality of honey is very crucial because pollen from poisonous plants can sometimes be found within honey, for example, *Atropa belladonna* (Bruni et al. 2015). Hepatotoxic pyrrolizidine alkaloids (PAs) have been detected in honeys after bees have foraged on plants within the Boraginaceae (Edgar et al. 2002) and grayanotoxins arising from *Rhododendron* spp. (Koca and Koca 2007). Thus botanical profile of honey is therefore very important to ensure that products are of high quality and safe for the user (Olivieri et al. 2012).

Food Labeling, Food Safety, and Food Piracy In food industry and food supply chains, cpDNA and mtDNA barcoding are used for food labeling, food safety, and food piracy in processed products. Since several copies of extranuclear genomes are within each cell, this technique helps to identify the information of degraded samples or transformed materials derived from crops and livestock species. The technology is used for routine analyses to maintain food safety and quality. DNA barcoding can be

crucial because it can verify the presence/absence of the original species and to identify the nature of the replaced species. One of the most striking substitution cases ever revealed refers to fish meat (e.g., sold as slices, fillets, blocks, surimi, fish sticks, and fins).

Adulterants in foods are or adding any substance to increase weight and lose its original nature and appears to be better than original products. Most adulterants are economically motivated misbranding and mislabeling, fakes based on simulation processes and imitation products. When the adulterants are toxic or allergenic, serious public health consequences may result. In this case, the food mislabeling not only cheats the users, but it also causes intolerance or allergies to certain foods or their components. The most frequent incidents as per the literature from 1980 were grouped into 11 food categories: fish and seafood, dairy products, fruit juices, oils and fats, grain products, honey and other natural sweeteners, spices and extracts, wine and other alcoholic beverages, infant formula, plant-based proteins, and other food products (Barcaccia et al. 2015).

5.2.4 Forensics

Forensic palynology is the use of pollen to link persons or objects which related to particular place and time (Mathewes 2006). This technique is of great utility to forensics because (1) pollen is a nearly ubiquitous feature of the environment; (2) different geographic locations have different pollen signatures, allowing for inference related to spatial tracking; (3) plants flower at different times, allowing for temporal inference; and (4) pollen is extremely durable and thus the sample can be studied any time (Walsh and Horrocks 2008). DNA metabarcoding application could very likely increase its usage at broader range of situations (Bell et al. 2016). For example, on combining DNA barcoding with a universal database of geographic and temporal knowledge of plants (Goodman et al. 2015), the database could potentially be useful in forensics. In any kind of genetic analysis, DNA barcoding requires damaged pollen grain samples; this issue can be addressed to split pollen samples into partitions for DNA extraction, morphological examination, and permanent storage. The results of the DNA barcoding and morphological identifications are compared for its consistency and accuracy. In the future DNA metabarcoding will use pollen as a biomarker, which provides forensic scientists to ensure global security and justice.

5.2.5 Biomedicine

Metabarcoding is used to detect species in complex traditional Chinese medicine (TCM) samples presented in various forms (powders, crystals, capsules, tablets, and herbal tea). Screening of these samples reveals that the samples contained CITES-listed species, including the Asiatic black bear (*Ursus thibetanus*) and the Saiga antelope (*Saiga tatarica*), as well as unlisted ingredients, and potentially toxic and

allergenic plants (Coghlan et al. 2012). Metabarcoding analyses on standard Chinese medicines based on a six-herb formula found that significant differences in quality and safety of commercial TCM preparations, since the identification of species *Senna obtusifolia has* potential risks to consumers. Combination of NGS and DNA barcoding successfully applied in species identity in herbal products (Roth et al. 2016). To identify medicinal plants, *matK, rbcK, trnH-psbA, ITS, trnL-F, 5S-rRNA and 18S-rRNA* genes are used.

5.2.6 Agriculture Protection

In Agriculture, to control pests, DNA barcoding helps in identifying pests in any stage of life, this method is used by the farmers to save agricultural crops from the pest damage. The global tephritid barcoding initiative contributes to identify, manage and control fruit flies. DNA barcoding is useful for rapid, accurate, and cost-effective in identifying specimens and determining whether they are native, invasive or should be quarantined. For farmers, DNA barcoding facilitates the rapid identification of bugs pests or non-pests in the field to take preventive measures to control pests. It also offers creation accurate catalogue of life on earth; which can be used to monitor biodiversity for changes and secure ecosystems. Through NGS technologies, DNA sequencing and macroarray diagnostics protocols that greatly reduce the impact of harmful plant pathogens in the field and environment (Agriculture and Agri-Food Canada 2015). The accuracy and efficiency of DNA-based species identification makes DNA barcoding is very useful in vector surveillance and bio-security programs (Batovska et al. 2016).

6 Conclusions

DNA barcoding is a useful molecular genetic-based technique in terms of identifying unknown species or specimens. DNA barcoding helps to differentiate similar-looking species and identify immature or damaged specimens. Although several gene markers have been suggested, more intensive work is needed to verify whether DNA barcodes can successfully provide accurate identification across all taxa and species with hybridization. DNA metabarcoding has been applied in next-generation sequencing technologies in genomic analysis of complex environmental samples and is becoming an important tool for understanding evolutionary history and ecological biodiversity. Building a well-covered DNA barcoding reference library is central to use as benchmark for validation and successful identification of species biodiversity. Regulatory authorities are benefiting from barcoding technology to test specimens and precisely identify trade commodities. In fundamental research, DNA barcodes are enormously and expandingly used in taxonomy, ecology, evolution, and biogeography. In applied research, DNA barcodes are offerening some of the

most exciting prospects to be used to monitor and control human diseases and manage agricultural pathogens and biosecurity and quarantine issues.

Acknowledgments SD acknowledges the funding of UGC-RGNF, New Delhi. BRN gratefully acknowledges funding of DSTPURSE programme. The research in the laboratory of AS supported by DST-FIST, UGC-CAS and DST-PURSE, New Delhi are gratefully acknowledged.

References

Agriculture and Agri-Food Canada (2015) DNA barcoding – science helping farmers identify friend from foe. Media Relations

Anklam E (1998) A review of the analytical methods to determine the geographical and botanical origin of honey. Food Chem 63(4):549–562

Baird N, Etter P, Atwood T et al (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3:e3376

Barcaccia G, Lucchin M, Cassandro M (2015) DNA barcoding as a molecular tool to track down mislabeling and food piracy. Diversity 8(1):2

Batovska J, Blacket MJ, Brown K, Lynch SE (2016) Molecular identification of mosquitoes (Diptera: Culicidae) in southeastern Australia. Ecol Evol 9(9):3001–3011

Bell KL, Burgess KS, Okamoto KC, Aranda R, Brosi BJ (2016) Review and future prospects for DNA barcoding methods in forensic palynology. Forensic Sci Int Genet 21:110–116

Besnard G, Christin P-A, Malé P-JG et al (2014) From museums to genomics: old herbarium specimens shed light on a C3 to C4 transition. J Exp Bot 65:6711–6721

Briskin DP (2000) Medicinal plants and phytomedicines. Linking plant biochemistry and physiology to human health. Plant Physiol 124:507–514

Brosi BJ, Daily GC, Shih TM, Oviedo F, Durán G (2007) The effects of forest fragmentation on bee communities in tropical countryside. J Appl Ecol 45:773–783

Bruni I, De Mattia F, Galimberti A, Galasso G, Banfi E, Casiraghi M, Labra M (2010) Identification of poisonous plants by DNA barcoding approach. Int J Legal Med 124:595–603

Bruni I, Galimberti A, Caridi L, Scaccabarozzi D, De Mattia F, Casiraghi M, Labra M (2015) A DNA barcoding approach to identify plant species in multiflower honey. Food Chem 170:308–315

Burgess KS, Fazekas AJ, Kesanakurti PR, Graham SW, Husband BC, Newmaster SG, Percy DM, Hajibabaei M, Barrett SCH (2011) Discriminating plant species in a local temperate flora using the rbcL+matK DNA barcode. Methods Ecol Evol 2:333–340

Coghlan ML, Haile J, Houston J, Murray DC, White NE, Moolhuijzen P, Bellgard MI, Bunce M (2012) Deep sequencing of plant and animal DNA contained within traditional Chinese medicines reveals legality issues and health safety concerns. PLoS Genet 8(4):e1002657. https://doi.org/10.1371/journal.pgen.1002657

Cook CN, Mascia MB, Schwartz MW, Possingham HP, Fuller RA (2013) Achieving conservation science that bridges the knowledge-action boundary. Conserv Biol 27:669–678

Crane E (1975) Honey. A comprehensive survey. William Heinemann, London

D'Amato G, Liccardi G, D'Amato M, Holgate S (2005) Environmental risk factors and allergic bronchial asthma. Clin Exp Allergy 35(9):1113–1124

Davies JM, Beggs PJ, Medek DE, Newnham RM, Erbas B, Thibaudon M, Katelaris CH, Haberle SG, Newbigin EJ, Huete AR (2015) Trans-disciplinary research in synthesis of grass pollen aerobiology and its importance for respiratory health in Australasia. Sci Total Environ 534:85–96

De Mattia F, Gentili R, Bruni I, Galimberti A, Sgorbati S et al (2012) A multi-marker DNA barcoding approach to save time and resources in vegetation surveys. Bot J Linn Soc 169:518–529

DeSalle R (2006) Species discovery versus species identification in DNA barcoding efforts: response to Rubinoff. Conserv Biol 20:1545–1547
Dunning LT, Savolainen V (2010) Broad-scale amplification of matK for DNA barcoding plants, a technical note. Bot J Linn Soc 164:1–9
Edgar B, Doherty J, Meert H (2002) Access to housing, homelessness and vulnerability in Europe. Policy Press, the JCSHR and FEANTSA, Bristol
Falade MO, Opene AJ, Benson O (2016) DNA barcoding of Clarias gariepinus, Coptodon zillii and Sarotherodon melanotheron from Southwestern Nigeria. F1000Res eCollection
Fan H, Ives AR, Surge-groba Y, Cannon AH (2015) An assembly and alignment-free method of phylogeny reconstruction from next-generation sequencing data. BMC Genomics 16:522
Fazekas AJ, Kesanakurti PR, Burgess KS, Percy DM, Graham SW, Barrett SC, Newmaster SG, Hajibabaei M, Husband BC (2009) Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? Mol Ecol Resour 9:130–139
Ficetola GF, Miaud C, Pompanon F, Taberlet P (2008) Species detection using environmental DNA from water samples. Biol Lett 4:423–425
Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW, Delong EF (2008) Microbial community gene expression in ocean surface waters. Proc Natl Acad Sci USA 105 (10):3805–3810
Galimberti A, De Mattia F, Bruni I, Scaccabarozzi D, Sandionigi A, Barbuto M, Casiraghi M, Labra M (2014) A DNA barcoding approach to characterize pollen collected by honeybees. PLoS One 9(10):e109363. https://doi.org/10.1371/journal.pone.0109363
Goodman FJ, Doughty JW, Gary C, Christou CT, Hu BB, Hultman EA, et al (2015) PIGLT: a pollen identification and geolocation system for forensic applications. In: 2015 I.E. international symposium on Technologies for Homeland Security (HST). pp 1–7
Hajibabaei M, Shokralla S, Zhou X, Singer GA, Baird DJ (2011) Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS One 6:e17497
Hawkins J, de Vere N, Griffith A, Ford CR, Allainguillaume J, Hegarty MJ, Baillie L, Adams-Groom B (2015) Using DNA metabarcoding to identify the floral composition of honey: a new tool for investigating honey bee foraging preferences. PLoS One 10(8):e0134735. https://doi.org/10.1371/journal.pone.0134735
Hebert PD, Cywinska A, Ball SL (2003a) Biological identifications through DNA barcodes. Proc R Soc Lond B Biol Sci 270(1512):313–321
Hebert PD, Ratnasingham S, de Waard JR (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc R Soc Lond B Biol Sci 270(Suppl 1): S96–S99
Hegland SJ, Nielsen A, Lázaro A, Bjerknes A-L, Totland Ø (2009) How does climate warming affect plant-pollinator interactions? Ecol Lett 12:184–195
Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Mol Ecol Resour 11:117–122
Hollingsworth ML, Clark A, Forrest LL, Richardson JR, Pennington RT et al (2009) Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. Mol Ecol Resour 9:439–457
Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. PLoS One 6:e19254
Huber JA et al (2007) Microbial population structures in the deep marine biosphere. Science 318:97–100
http://www.barcodeoflife.org/content/about/what-cbol. The Consortium for the Barcode of Life (CBOL)
Inouye DW (2008) Effects of climate change on phenology, frost damage, and floral abundance of montane wildflowers. Ecology 89:353–362
Janda JM, Abbott SL (2007) 16SrRNA gene sequencing for bacterial identification in the diagnostic laboratory pluses, perils and pitfalls. J Clin Microbiol 45:2761–2764

Jones YL, Peters SM, Weland C, Ivanova NV, Yancy HF (2013) Potential use of DNA barcodes in regulatory science: identification of the US food and drug administrations "Dirty 22" contributors to the spread of foodborne pathogens. J Food Prot 76:144–149

Jordano P (2010) Pollen, seeds and genes: the movement ecology of plants. Heredity 105:329–330

Jorgensen T, Haile J, Moller P, Andreev A, Boessenkool S, Rasmussen M et al (2012a) A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals longterm vegetational stability. Mol Ecol 21(8):1989–2003

Jorgensen T, Kjaer KH, Haile J, Rasmussen M, Boessenkool S, Andersen K et al (2012b) Islands in the ice: detecting past vegetation on Greenlandic nunataks using historical records and sedimentary ancient DNA metabarcoding. Mol Ecol 21(8):1980–1988

Kane NC, Cronk Q (2008) Botany without borders: barcoding in focus. Mol Ecol 17:5175–5176

Keele J, Carmon J, Pucherelli SF, Hosler D (2014) Identification of unknown organisms by DNA barcoding: a molecular method for species classification research and development office invasive mussels. Final report 2014-01 (0045)

Keller A, Danner N, Grimmer G, Ankenbrand M, von der Ohe K, von der Ohe W, Rost S, Härtel S, Steffan-Dewenter I (2015) Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. Plant Biol 17(2):558–566

Koca I, Koca AF (2007) Poisoning by mad honey: a brief review. Food Chem Toxicol 45 (8):1315–1318

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol 79(17):5112–5120

Kraaijeveld K, de Weger LA, Ventayol García M, Buermans H, Frank J, Hiemstra PS, den Dunnen JT (2015) Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. Mol Ecol Resour 15(1):8–16

Kress WJ, Carlos G-R, Uriarte M, Erickson DL (2015) DNA barcodes for ecology, evolution, and conservation. Trends Ecol Evol 30(1):25–35

Laube I, Hird H, Brodmann P, Ullmann S, Schöne-Michling M, Chisholm J, Broll H (2010) Development of primer and probe sets for the detection of plant species in honey. Food Chem 118(4):979–986

Lebonah DE, Dileep A, Chandrasekhar K, Sreevani S, Sreedevi B, Pramoda Kumari J (2014) DNA barcoding on bacteria: a review. Article ID 541787, 9 pages

Li D-Z et al (2011) Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proc Natl Acad Sci U S A 108:19641–19646

Links MG, Dumonceaux TJ, Hemmingsen SM, Hill JE (2012) The Chaperonin-60 universal target is a barcode for bacteria that enables de novo assembly of metagenomic sequence data. PLoS One 7(11):e49755

Little DP (2014) A DNA mini-barcode for land plants. Mol Ecol Resour 14:437–446

Liu S-YV, Chan C-LC, Lin O, Hu C-S, Chen CA (2013) DNA barcoding of shark meats identify species composition and CITES-listed species from the markets in Taiwan. PLoS One 8(11): e79373

Machida RJ, Hashiguchi Y, Nishida M, Nishid S (2009) Zooplankton diversity analysis through single-gene sequencing of a community sample. BMC Genomics 10(1):438

Maillet N, Collet G, Vannier T, Lavenier D, Peterlongo P (2014) Commet: comparing and combining multiple metagenomic datasets. In: IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp 94–98

Mathewes RW (2006) Forensic palynology in Canada: an overview with emphasis on archaeology and anthropology. Forensic Sci Int 163:198–203

Matsen FA, Kodner RB, Armbrust EV (2010) pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics 11:538

McCormack B, Rycroft-Malone J, DeCorby K, Hutchinson AM, Bucknall T, Kent B, Schultz A, Snelgrove-Clarke E, Stetler C, Titler M, Wallin L, Wilson V (2013) A realist review of interventions and strategies to promote evidence-informed healthcare: a focus on change agency. Implement Sci 8:107

Medinger R, Nolte V, Pandey RV, Jost S, Ottenwälder B, Schlötterer C, Boenigk J (2010) Diversity in a hidden world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. Mol Ecol 19(1):32–40

Mou X, Sun S, Edwards RA, Hodson RE, Moran MA (2008) Bacterial carbon processing by generalist species in the coastal ocean. Nature 451:708–711

Muturi CN, Ouma JO, Malele II, Ngure RM, Rutto JJ, Mithöfer KM, Enyaru J, Masiga DK (2011) Tracking the feeding patterns of tsetse flies (*Glossina* genus) by analysis of bloodmeals using mitochondrial cytochromes genes. PLoS One 6:e17284

Natalie JWW (2013) Determining microbial diversity in chaparral soils before and after wildfires through DNA barcoding Project No. S-1119 California State Science Fair Project Summary

Newmaster SG, Ragupathy S, Janovec J (2009) A botanical renaissance: state-of-the-art DNA bar coding facilitates an automated identification technology system for plants. Int J Comput Appl Technol 35:50–60

Newsmaster SG, Grguric M, Shanmughanandhan D, Ramalingam S, Ragupathy S (2013) DNA barcoding detects contamination and substitution in North American herbal products. BMC Med 11:1–13

Olivieri C, Marota I, Rollo F, Luciani S (2012) Tracking plant, fungal, and bacterial dna in honey specimens. J Forensic Sci 57(1):222–227

Parducci L, Matetovici I, Fontana SL, Bennett KD, Suyama Y, Haile J, Kjaer KH, Larsen NK, Drouzas AD, Willerslev E (2013) Molecular- and pollen-based vegetation analysis in lake sediments from central Scandinavia. Mol Ecol 22(13):3511–3524

Pavan-Kumar A, Gireesh-Babu P, Lakra WS (2015) DNA metabarcoding: a new approach for rapid biodiversity assessment. J Cell Sci Mol Biol 2(1):111

Pedersen MW, Ginolhac A, Orlando L, Olsen J, Andersen K, Holm J, Funder S, Willerslev E, Kjær KH (2013) A comparative study of ancient environmental DNA to pollen and macrofossils from lake sediments reveals taxonomic overlap and additional plant taxa. Quat Sci Rev 75:161–168

Pennisi E (2000) Taxonomic revival. Science 289:5488

Persano Oddo L, Bogdanov S (2004) Determination of honey botanical origin: problems and issues. Apidologie 35(Suppl 1):S2–S3

Primack R, Rozzi R, Feinsinger P, Dirzo R, Massardo F (2001) Fundamentos de conservación biológica: Perspectivas Latinoamericanas. Fondo de Cultura Economica México D.F.797

Ratnasingham S, Hebert PD (2007) bold: The Barcode of Life Data System (www.barcodinglife.org). Mol Ecol Notes 7:355–364

Razgour O, Clare EL, Zeale MRK, Hanmer J, Schnell IB, Rasmussen M, Gilbert TP, Jones G (2011) High-throughput sequencing offers insight into mechanisms of resource partitioning in cryptic bat species. Ecol Evol 1(4):556–570

Richardson RT, Lin C-H, Quijia JO, Riusech NS, Goodell K, Johnson RM (2015a) Rank-based characterization of pollen assemblages collected by honey bees using a multi-locus metabarcoding approach. Appl Plant Sci 3(11):1500043

Richardson RT, Lin C-H, Sponsler DB, Quijia JO, Goodell K, Johnson RM (2015b) Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. Appl Plant Sci 3(1):1400066

Ricketts TH, Regetz J, Steffan-Dewenter I, Cunningham SA, Kremen C, Bogdanski A, Gemmill-Herren B, Greenleaf SS, Klein AM, Mayfield MM, Morandin LA, Ochieng' A, Viana BF (2008) Landscape effects on crop pollination services: are there general patterns? Ecol Lett 11:499–515

Roth F, Yachie N, Petsalaki E, Mellor JC, Weile J, Jacob Y, Verby M, Ozturk SB, Li S, Cote AG et al (2016) Pooled-matrix protein interaction screens using barcode fusion genetics. Mol Syst Biol 12:863

Sahare P, Srinivasu T (2012) Barcoding for authentic identification of medicinal plants. Int J Eng Sci 1:33–36

Scheifinger H, Belmonte J, Buters J, Celenk S, Damialis A, Dechamp C, García-Mozo H, Gehrig R, Grewling L, Halley JM, Hogda K-A, Jäger S, Karatzas K, Karlsen S-R, Koch E, Pauling A, Peel R, Sikoparija B, Smith M, Galán-Soldevilla C, Thibaudon M, Vokou D, de Weger LA (2013) Monitoring, modelling and forecasting of the pollen season. Allergenic Pollen 71–126. Springer Netherlands, Dordrecht.

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, JL ALC, Chen W (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. Proc Natl Acad Sci USA 109:6241–6246

Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am J Bot 94(3):275–288

Shokralla S, Gibson JF, Nikbakht H, Janzen DH, Hallwachs W, Hajibabaei M (2014) Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. Mol Ecol Resour 14(5):892–901

Sickel W, Ankenbrand MJ, Grimmer G, Holzschuh A, Härtel S, Lanzen J, SteffanDewenter I, Keller A (2015) Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. BMC Ecol 15(1):20

Smith MA, Bertrand C, Crosby K, Eveleigh ES, Fernandez-Triana J, Fisher BL et al (2012) Wolbachia and DNA barcoding insects: patterns, potential and problems. PLoS One 7:e36514

Sogin ML, Morrison HG, Huber JA et al (2006) Microbial diversity in the deep sea and the underexplored 'rare biosphere'. Proc Natl Acad Sci USA 103(32):12115–12120

Stoeck T, Bass D, Nebel M, Christen R, Jones MD et al (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Mol Ecol 19(Suppl 1):21–31

Suyama Y, Kawamuro K, Kinoshita I, Yoshimura K, Tsumura Y, Takahara H (1996) DNA sequence from a fossil pollen of Abier spp. from Pleistocence peat. Genes Genet Syst 7 (1):145–149

Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G, Brochmann C, Willerslev E (2007) Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Res 35:e14–e14

Thao NP, Thu NB, Hoang XL, Van Ha C, Tran LS (2013) Differential expression analysis of a subset of drought-responsive *GmNAC* genes in two soybean cultivars differing in drought tolerance. Int J Mol Sci 14:23828–23841

Valentini A, Miquel C, Nawaz MA, Bellemain E, Coissac E, Pompanon F, Gielly L, Cruaud C, Nascetti G, Wincker P, Swenson JE, Taberlet P (2009) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. Mol Ecol Resour 9 (1):51–60

Valentini A, Miquel C, Taberlet P (2010) DNA barcoding for honey biodiversity. Diversity 2:610–617

Vivas CV, Moraes RCS, Alves-Araújo A, Alves M, Mariano-Neto E, van den Berg C, Gaiotto FA (2014) DNA barcoding in Atlantic Forest plants: what is the best marker for Sapotaceae species identification? Genet Mol Biol 37(4):662–670

Vohra P, Khera KS (2013) DNA barcoding: current advance and future prospects—a review. Asian J Biol Life Sci 3:185–189

Walsh KA, Horrocks M (2008) Palynology: its position in the field of forensic science. J Forensic Sci 53(5):1053–1060

Williamson SJ, Rusch DB, Yooseph S, Halpern A, Heidelberg KB, Glass JI et al (2008) The sorcerer II global ocean sampling expedition: metagenomic characterization of viruses within aquatic microbial samples. PLoS One 1:e1456

Wilson EE, Sidhu CS, LeVan KE, Holway DA (2010) Pollen foraging behavior of solitary Hawaiian bees revealed through molecular pollen analysis. Mol Ecol 19:4823–4829

Wilson JJ, Sing KW, Halim MR, Ramli R, Hashim R, Sofian-Azirun M (2014) Utility of DNA barcoding for rapid and accurate assessment of bat diversity in Malaysia in the absence of formally described species. Genet Mol Res 13(1):920–925

Zhang AB, Muster C, Liang HB, Zhu CD, Crozier R, Wan P et al (2012) A fuzzy-set-theory-based approach to analyse species membership in DNA barcoding. Mol Ecol 21:1848–1863

R in DNA Barcoding

Asim Kumar Mahadani, Pradosh Mahadani, and Goutam Sanyal

Abstract DNA barcoding bloomed as a routine application in biodiversity assessment and species identification. Several improvements have occurred in the DNA sequence analysis approaches. Recently, R language and environment for statistical computing and graphics emerged as powerful tool in sequence and phylogenetic analysis. Bioinformaticians have written several specialized packages for R to provide reading and writing data and manipulating sequence analysis, as well as several advanced DNA barcode sequence-based analysis. SeqinR package is mainly used to retrieve and analyze biological sequences, and APE packages are for construction of phylogenetic and evolutionary tree-based molecular data. However, SPIDER and BarcodingR packages are completely devoted to analyze for DNA barcode sequences. The purpose of this chapter is to provide a quick reference for DNA barcode associate researcher in sequence and phylogenetic analysis.

Keywords R language · DNA barcoding · APE · SPIDER · BarcodingR

1 Introduction

DNA barcoding aims to assign each individual to respective species according to the standardized short genome sequences. Species identification using DNA barcode sequences routinely applied in biodiversity studies of animal, plant, and fungi and accumulate large amount of DNA sequences. BOLD system (www.barcodinglife.org) is a comprehensive reference library of DNA barcode sequences (Ratnasingham and Hebert 2007). However, very few user-friendly and noncommercial software are available to analyze the large amount of DNA barcode sequences in single platform. Recently, R language and environment for statistical computing and graphics

A. K. Mahadani · G. Sanyal (✉)
Department of Computer Science & Engineering, National Institute of Technology, Durgapur, West Bengal, India

P. Mahadani
Crop Improvement Division, ICAR-National Rice Research Institute, Cuttack, India

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_2

emerged as powerful tool in DNA barcode sequence interpretation. It is free and open-source software, licensed under the GNU General Public License, with copyright held by "the R Foundation for Statistical Computing." R is created by Ross Ihaka and Robert Gentleman at the University of Auckland. The main advantages of "R" are freeware, outstanding graphical capabilities. There is a large amount of (free) documentation and help available from online. Some help is also already installed along with the packages. To get help from R console, type help(). It will show how to use help() function to get the documentation of any function or method. R can do many statistical and data analyses (R Development Core Team 2011) and organized in the so-called packages or libraries. With the standard installation, most common packages are installed. We can manually strike it to load library. There are many more packages available on the R website. R is an interpreter that uses a command-line-based environment. This indicates that we have to type commands, rather than use the mouse and menus. The advantage is that we do not always have to retype all commands. We can store our commands in file which is known as scripts. These scripts have filenames typically with the extension .R. We can open an editor window to edit these files by clicking File and New or Open file. We can run (send to the console window) part of the code by selecting lines and pressing CTRL+ENTER or click Run in the editor window.

- Install the package: click install in the package window and type mclust or type install.packages("mclust") in the command window, or from the menu bar, we can select Tools>install packages and then specify mclust as package to be installed.
- Load the package: check box in front of mclust or type library ("mclust") in the command window.

This chapter covers basics of R software to carry out DNA sequence analysis as well as specialized R packages such as SeqinR, Analyses of Phylogenetics and Evolution (APE), Species Identity and Evolution in R (SPIDER), and BarcodingR for DNA barcoding studies.

2 SeqinR

The SeqinR is one of the primary packages to retrieve and analyze biological sequences (Charif and Lobry 2007). This package mainly retrieves sequences and is also used for visualization of biological sequence (DNA and protein) data. It includes also utilities for sequence data management under the ACNUC system (Table 1).

Table 1 List of important SeqinR functions with descriptions used in DNA barcode sequence retrieval, visualization, and analysis

Function	Description
aaa()	Converts amino acid one-letter code into the three-letter one, for instance, "A" into "Ala"
acnucopen()	Open and close a remote access to an ACNUC database
anoukResult()	Expected numeric results for Ka (substitutions per nonsynonymous site) and Ks (substitutions per synonymous site) computation between two protein-coding genes. They are also denoted as ds and dn in the literature. The ratio of nonsynonymous (Ka) to synonymous (Ks) nucleotide substitution rates is an indicator of selective pressures on genes
AAstat()	This function returns simple protein sequence information including the number of residues, the percentage physicochemical classes, and the theoretical isoelectric point
cai()	CAI stands for the Codon Adaptation Index that is used for the measurement of the relative adaptiveness of the codon usage of a gene toward the codon usage of highly expressed genes
choosebank()	To select a database under ACNUC and located on the web
as.matrix.alignment()	Converts an alignment into a matrix of characters
comp()	Complements a sequence, for instance, if the sequence is "a," "c," "g," and "t," it returns "t," "g," "c," and "a." This is not the reverse complementary strand
consensus()	Construct a consensus and profiles for sequence alignments
computePI()	This function calculates the theoretical isoelectric point of a protein. Isoelectric point is the pH at which the protein has a neutral charge
closebank()	To close a remote ACNUC database
dist.alignment()	These functions compute a matrix of pairwise distances from aligned sequences using similarity (Fitch matrix, for protein sequences only) or identity matrix (for protein and DNA sequences). The resulting matrix contains the squared root of the pairwise distances. This function is important for intra- and interspecies distance calculation in DNA barcode sequence analysis
gb2fasta	Converts a single entry in GenBank format into a fasta file
extractseqs()	To extract the sequence information of a sequence or a list of sequence in different formats
G+C Content()	Calculates the fractional G+C content of nucleic acid sequences. It reads nucleic acid sequences, sums the number of "g" and "c" bases, and writes out the result as the fraction (in the interval 0.0 to 1.0) to the total number of "a," "c," "g," and "t" bases
read.fasta()	Read nucleic or amino acid sequences from a file in fasta format. By default read.fasta return a list of vector of chars
Kaks ()	It computes the substitutions per synonymous site and nonsynonymous site
pK	pK values for the side chain of charged amino acids from various sources
read.alignment()	Read aligned sequence files in mase, clustal, phylip, fasta, or msf format
write.fasta()	Write sequence(s) into a file in fasta format

3 Reading DNA Sequence Data

The SeqinR package in R, can easily be used to read a DNA sequence from a fasta file, for example, how to read the COX1 DNA sequence of *Bos indicus* that collected from the NCBI database and save it in a fasta format file (e.g., "cox1.fasta"). We can read this fasta format file into R using the read.fasta() function.

After installing the SeqinR package in R, we can load the package and read the file using the follwing commands:

```
> library("seqinr")
> barcode <- read.fasta(file = "cox1.fasta ")
```

Note that R expects the files that we read in (e.g., "cox1.fasta") to be in the current working directory on our computer, so if we have stored "cox1.fasta" somewhere else, we shall have to move or copy it into the current working directory. The command above reads the contents of the fasta format file cox1.fasta into an R object called barcode. The variable barcode is an R list object. A list is an R object that is like a vector but can contain elements that are numeric and/or contain characters. In this case, the list barcode contains information from the fasta file that we have read in (i.e., the NCBI accession for the barcode sequence and the DNA sequence itself). In fact, the first element of the list object barcode contains the DNA sequence. We can access the elements of an R list object using double square brackets. Thus, we can store the DNA sequence for COX1 in a variable barcodeseq by typing:

```
> barcodeseq <- barcode[[1]]
```

The variable barcodeseq is a vector containing the nucleotide sequence. Each element of the vector contains one nucleotide of the sequence. Therefore, to print out a certain subsequence of the sequence, we just need to type the name of the vector barcodeseq followed by the square brackets containing the indices for those nucleotides. For example, the following command prints out the first 25 nucleotides of the COX 1 sequence:

```
> barcodeseq [1:25]
[1] "a" "g" "t" "t" "g" "t" "t" "a" "g" "t"
[11] "c" "t" "a" "c" "g" "t" "g" "g" "a""a"
[21] "c" "c" "g" "a" "c"
```

Note that barcodeseq[1:25] refers to the elements of the vector barcodeseq with indices from 1 to 25. These elements contain the first 25 nucleotides of the COX1 sequence.

4 Calculate the Length of a DNA Sequence

Once we have retrieved a DNA sequence, we can obtain some simple statistics to describe that sequence, such as the sequence's total length in nucleotides. To subsequently obtain the length of the DNA sequence, we would use the length() function, typing:

```
> length(barcodeseq)
[1] 685
```

The length() function will give us back the length of the sequence stored in variable barcodeseq, in nucleotides. The length() function actually gives the number of elements in the input vector that we pass to it, which in this case is the number of elements in the vector barcodeseq.

5 Determination of Base Composition of a DNA Sequence

An obvious first analysis of any DNA sequence is to count the number of occurrences of the four different nucleotides ("A," "C," "G," and "T") in the sequence. This can be done using the table() function. For example, to find the number of As, Cs, Gs, and Ts in the COX1 DNA sequence of *Bos indicus* (which we have put into vector variable barcodeseq, using the commands above), we should type:

```
> table(barcodeseq)
barcodeseq
  a   c   g   t
125 240 170 150
```

This means that the COX1 DNA sequence of *Bos indicus* has 125 As, 240 Cs, 170 Gs, and 150 Ts.

6 Calculation of GC Content of DNA Sequence

One of the most fundamental properties of a genome sequence is its GC content, the fraction of the sequence that consists of Gs and Cs, i.e., the %(G+C). The GC content can be calculated as the percentage of the bases in the genome that are Gs or Cs. That is, GC content = (number of Gs + number of Cs)*100/(genome length). For example, if the genome is 100 bp, and 20 bases are Gs and 21 bases are Cs, then the GC content is (20 + 21)*100/100 = 41%. We can easily calculate the GC content based on the number of As, Gs, Cs, and Ts in the genome sequence. For example, for the COX1

DNA sequence of *Bos indicus*, we know from using the table() function above that the genome contains 125 As, 240 Cs, 170 Gs, and 150 Ts.

Therefore, we can calculate the GC content using the command:

```
> (240+170)*100/(125+240+170+150)
[1] 59.8540
```

Alternatively, we can use the GC() function in the SeqinR package, which gives the fraction of bases in the sequence that are Gs or Cs.

```
> GC(barcodeseq)
[1] 0.598540
```

The result above means that the fraction of bases in the COX1 DNA sequence that are Gs or Cs is 0.4666977. To convert the fraction to a percentage, we have to multiply by 100, so the GC content as a percentage is 59.8540%.

7 DNA Words

As well as the frequency of each of the individual nucleotides ("A," "G," "T," "C") in a DNA sequence, it is also interesting to know the frequency of longer DNA "words." The individual nucleotides are DNA words that are one nucleotide long, but we may also want to find out the frequency of DNA words that are two nucleotides long (i.e., "AA," "AG," "AC," "AT," "CA," "CG," "CC," "CT," "GA," "GG," "GC," "GT," "TA," "TG," "TC," and "TT"), three nucleotides long (e.g., "AAA," "AAT," "ACG," etc.), four nucleotides long, etc. To find the number of occurrences of DNA words of a particular length, we can use the count() function from the R SeqinR package. For example, to find the number of occurrences of DNA words that are one nucleotide long in the sequence barcodeseq, we type:> **count()**

As expected, this gives us the number of occurrences of the individual nucleotides. To find the number of occurrences of DNA words that are two nucleotides long, we type:

```
> count(barcodeseq, 2)
```

8 Analyses of Phylogenetics and Evolution (APE)

Analyses of Phylogenetics and Evolution (APE) (Paradis et al. 2004) provides utility functions for reading, writing, plotting, and manipulating phylogenetic trees, analyses of comparative data in a phylogenetic framework, ancestral character analyses, analyses of diversification and macroevolution, computing distances from DNA sequences, reading and writing nucleotide sequences as well as graphical exploration of phylogenetic data (alex, trex, kronoviz), estimation of absolute evolutionary rates and clocklike trees using mean path lengths and penalized likelihood, dating trees with non-contemporaneous sequences, translating DNA into AA sequences, and assessing sequence alignments. Phylogeny estimation can be done with the NJ, BIONJ, ME, MVR, SDM, and triangle methods, and several methods handling incomplete distance matrices (NJ*, BIONJ*, MVR*, and the corresponding triangle method). Some functions call external applications (PhyML, Clustal, T-Coffee, Muscle) whose results are returned into R (Table 2). Adegenet package is related to ape package (Jombart 2008).

9 Installation and Load of Package

First we have to install the packages and then load packages:

```
> install.packages("ape", dep = TRUE)
```

Then load the packages using:

```
> library(ape)
```

10 Reading the DNA Sequence Data

The data used in this tutorial are DNA sequences of *Citrus* downloaded from GenBank (http://www.ncbi.nlm.nih.gov/genbank/). Alignments have been realized beforehand using standard tools (Clustalw2 for basic alignment and Jalview for refining the results) (Thompson et al. 1997).

To read the DNA sequences into R, we use read.dna from the ape package:

```
> dna <- read.dna(file = "citrus.fasta", format = "fasta")
> dna
```

Thirty-one DNA sequences in binary format are stored in a matrix.

Table 2 List of important functions and descriptions of APE used in DNA barcode sequence-based phylogenetic analysis

Function	Description
ace()	This function is used for ancestral character calculation
base.freq()	Calculate the frequencies (absolute or relative) of the four DNA bases (adenine, cytosine, guanine, and thymidine) from a sample of sequences
BIONJ()	Construct the tree based on an improved version of the neighbor-joining algorithm
chronoMPL()	This function calculates the node ages of a tree using the mean path length method
chronopl()	Calculate the molecular dating tree nodes with penalized likelihood
Consensus()	The strict-consensus tree generates using this function
dist.dna()	This function computes a matrix of pairwise distances from DNA sequences using a model of DNA evolution. Eleven substitution models are currently available
del.gaps()	This function removes the insertion gaps ("-") in a sample of DNA sequences
DNAbin2indel()	This function scans a set of aligned DNA sequences and returns a matrix with information of the localizations and lengths on alignment gaps
nodepath()	This function searches all the paths from the root to each tip of the tree. If both arguments from and to are specified, the shortest path of nodes linking them is returned
nj()	Construct the neighbor-joining tree
read.tree()	This function reads the Newick or New Hampshire format
read.GenBank()	This function connects to the GenBank database and reads nucleotide sequences using accession numbers given as arguments
dist.gene()	This function computes a matrix of distances between pairs of individuals from a matrix or a data frame of genetic data
read.dna()	This function reads DNA sequences in a file and returns a matrix or a list of DNA sequences with the names of the taxa read in the file as rownames or names, respectively
speciesTree()	Estimate the species tree
summary.phylo()	This function gives summary of a phylogeny
Root()	Construct root phylogenetic trees
write.nexus.data()	Write character data in NEXUS format

All sequences are of same length: 446.

Labels: GQ267054, GQ267057, GQ267062, GQ435445, GQ435434, GQ248268 ....

Base composition:

```
  a   c   g   t
0.282 0.116 0.167 0.434
> class(dna)
[1] "DNAbin"
> as.character(dna)[1:8, 1:13]
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
```

```
GQ267054     "g" "g" "t" "c" "t" "t" "t" "g" "t" "g" "t" "a" "g"
GQ267057     "g" "g" "t" "c" "t" "t" "t" "g" "t" "g" "t" "a" "g"
GQ267062     "g" "g" "t" "c" "t" "t" "t" "g" "t" "g" "t" "a" "g"
GQ435445     "g" "g" "t" "c" "t" "t" "t" "g" "t" "g" "t" "a" "g"
GQ435434     "g" "g" "t" "t" "t" "t" "t" "g" "t" "g" "t" "a" "g"
GQ248268     "g" "g" "t" "c" "t" "t" "t" "g" "t" "g" "t" "a" "g"
EF590680     "g" "g" "t" "c" "t" "t" "t" "g" "t" "g" "t" "a" "g"
SGPM-MPT45   "g" "g" "t" "c" "t" "t" "t" "g" "t" "g" "t" "a" "g"
```

In R, different algorithms are available for constructing trees from a distance matrix including:

- nj (ape package): the classical neighbor-joining algorithm (Saitou and Nei 1987)
- bionj (ape): an improved version of neighbor-joining
- fastme.bal and fastme.ols (ape): minimum evolution algorithms
- hclust (stats): classical hierarchical clustering algorithms including single linkage, complete linkage, UPGMA, and others

Distance matrix calculation using dist.dna function: Here, we use Tamura and Nei 1993s model which considers different rates of transitions and transversions, heterogeneous base frequencies, and between site variation of the substitution rate.

```
> D <- dist.dna(dna, model = "TN93")
> class(D)
[1] "dist"
> tre <- nj(D)
> class(tre)
[1] "phylo"
> tre
```

Phylogenetic tree with 31 tips and 29 internal nodes.
Tip labels:

```
GQ267054, GQ267057, GQ267062, GQ435445, GQ435434, GQ248268, ....
```

Unrooted; includes branch lengths: We can plot NJ tree without bootstrap values (Fig. 1) with the following commands:

```
> plot(tre, show.tip = TRUE, edge.width = 2)
> title("NJ tree ")
> myBoots <- boot.phylo(tre, dna, function(e) root(nj(dist.dna(e,
model = "TN93")), 1))
```

Running bootstraps: 100/100.
Calculating bootstrap (Felsenstein 1981) values... done (Fig. 2). We can plot NJ tree with bootstrap values (Fig. 2) with the follwing commands:

```
> myBoots
 [1] 100  0  2  0  2  7  0  0  5  2  1 10  6 61 10  5  6 16 90 47 59 76 90 65
14 12  9 93 100
```

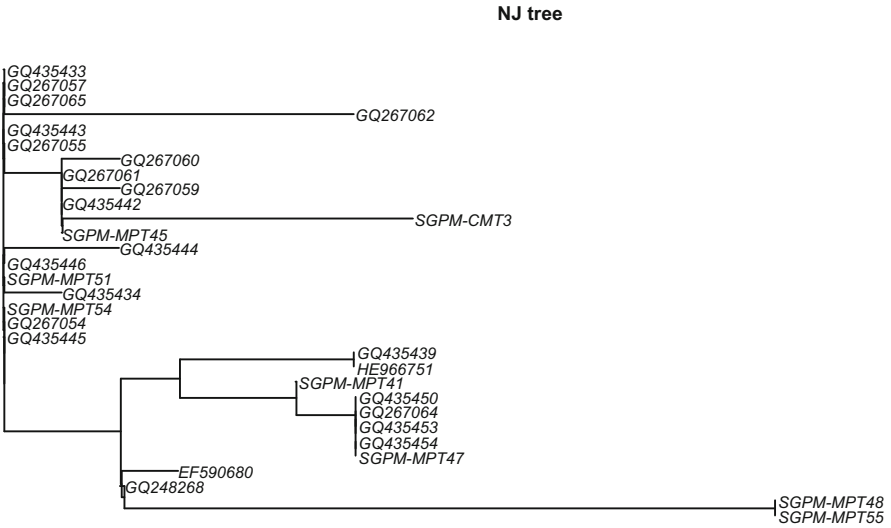


Fig. 1 Phylogenetic tree constructed using NJ method without bootstrap value

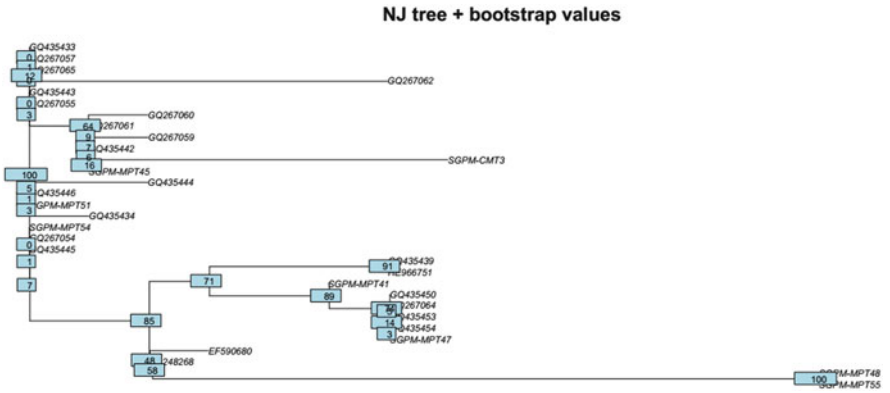


Fig. 2 Phylogenetic tree constructed using NJ method with bootstrap value

```
> plot(tre, show.tip = FALSE, edge.width = 2)
> title("NJ tree + bootstrap values")
>nodelabels(myBoots, cex = 0.6)
```

```
dm <- dist.ml(primates)
treeUPGMA <- upgma(dm)
```

We can plot the trees treeUPGMA (Fig. 3) with the commands:

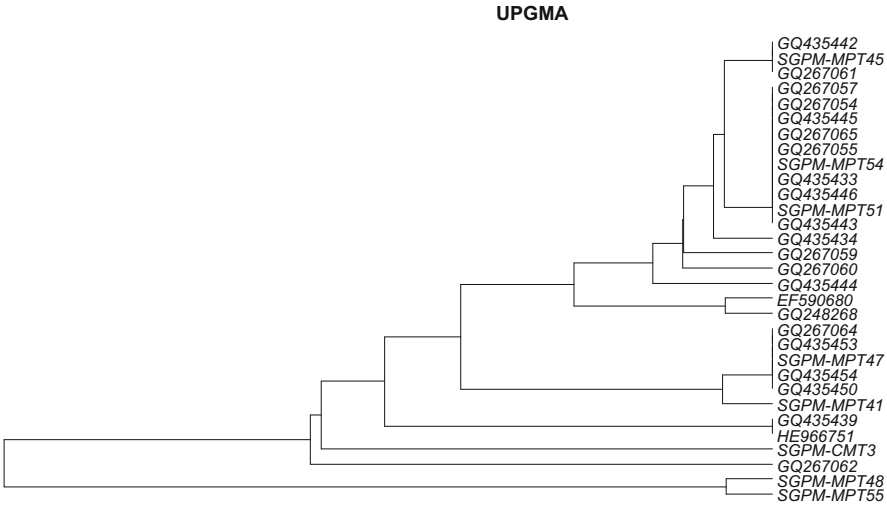


Fig. 3 Phylogenetic tree constructed using UPGMA method

```
layout(matrix(c(1,2), 2, 1), height=c(1,2))
par(mar = c(0,0,2,0) + 0.5)
plot(treeUPGMA, main="UPGMA")
```

11 Species Identity and Evolution in R (SPIDER)

Species Identity and Evolution in R (Brown et al. 2012) is an R package implementing a number of the most useful analyses for DNA barcoding studies and other research into species delimitation and speciation. It has been developed primarily by staff and students in the Department of Ecology at Lincoln University and was first released in 2011. The package seeks to provide a number of analyses that the developers have found useful in their own research and hadn't been implemented in other R packages (Table 3).

12 BarcodingR

Recently Zhang et al. (2016) developed an integrated BarcodingR package for species identification using DNA barcodes. BarcodingR is open-source for MICROSOFT Windows, MAC OSX, and Linux. BarcodingR have some additional function such as barcode evaluation, barcode gap analysis, delimitation comparison analysis, etc. which are user-friendly or noncommercial implementations (Table 4).

Table 3 List of important functions and their descriptions of SPIDER package used in DNA barcoding sequence analysis is given

Function	Description
seeBarcode()	It plots an illustrative barcode consisting of vertical bands in four colors corresponding to the DNA bases adenine (A), cytosine (C), guanine (G), and thiamine (T). Green, blue, black, and red are the standard colors representing A, G, C, and T, respectively. Plots an illustrative barcode
checkDNA()	Counts the number of bases in an alignment that are composed of missing data. This function considers bases coded as "?" and "N" as missing data. By default, gaps (coded as "-") are also considered missing. A numeric vector giving the number of missing bases in each sequence of the alignment
nucDiag()	Determines the diagnostic nucleotides for each species given in sppVector (the species vector). A list giving the pure, simple diagnostic nucleotides (i.e., those nucleotides that are fixed within species and different from all other species) for each species in the species vector. A result of integer(0) indicates there are no diagnostic nucleotides for those species
dataStat()	Check the numbers of species, genera, and individuals in the dataset. The value *NULL* can be passed to gen if genera are not of interest in the dataset. A table giving the number of genera and species in the dataset; giving the minimum, maximum, mean, and median number of individuals per species; and the number of species below the given threshold
nearNeighbour()	These functions test barcoding efficacy and are not identification tools. All sequences must be identified prior to testing. Each sequence is considered an unknown, while the remaining sequences in the dataset constitute the DNA barcoding database that is used for identification. If the identification from the test is the same as the pre-considered identification, a correct result is returned. bestCloseMatch conducts the "best close match" analysis, considering the closest individual unless it is further than the given threshold, which results in no identification. More than one species tied for closest match results in an assignment of "ambiguous." When the threshold is large, this analysis will return essentially the same result as nearNeighbour. nearNeighbour finds the closest individual and returns if their names are the same (*TRUE*) or different (*FALSE*). If names = *TRUE*, the name of the closest individual is returned. Ties are decided by majority rule. threshID conducts a threshold-based analysis, similar to that conducted by the "Identify Specimen" tool provided by the Barcode of Life Database (http://www.boldsystems.org/views/idrequest.php). It is more inclusive than bestCloseMatch, considering *ALL* sequences within the given threshold
seqStat()	This function considers bases coded as "?," "N," and "-" as missing data. A table giving the minimum, maximum, mean, and median sequence lengths and the number of sequences with lengths below the given threshold
monophyly()	Determines if the species given in sppVector form monophyletic groups on a given tree. monophyly determines if each species is monophyletic. monophylyBoot incorporates a bootstrap test to determine the support for this monophyly. Species with a bootstrap support lower than "thresh" are recorded as *FALSE*. Rerooting is done on the longest internal edge in the tree returned by nj(dist.dna(DNAbin))
read.BOLD()	Downloads DNA sequences from the Barcode of Life Database (BOLD)
rmSingletons ()	Detect and remove singletons

(continued)

Table 3 (continued)

Function	Description
sppDistMatrix()	Calculate the mean intra- and interspecific distance matrix
Titv()	Calculates the number of pairwise transitions and transversions between sequences
sppDist()	Calculate distance matrix based on its inter- and intraspecific components. There is option to produce histograms and other charts related to "barcode gap" measurement
read.GB()	Download sequences from GenBank with metadata. This function is a modification of read.GenBank() to include metadata with each sequence

Table 4 List of important functions and descriptions of BarcodingR used in phylogenetic analysis of DNA barcode

Function	Description
barcoding.gap()	Used for calculation of DNA barcoding gap. The results provide a list of summary statistics and plot distribution of interspecific and intra-specific genetic distance
barcodes.eval()	Test the efficiency of two barcodes using species identification success rate criteria
NAMES()	Extracts labels of samples
barcoding.spe.identify()	Species identification using protein-coding barcodes. Species identification using protein-coding barcodes with different methods, including BP-based method, fuzzy-set-based method, and Bayesian-based method
barcoding.spe.identify2()	Species identification using protein-coding barcodes with different methods, including BP-based method, fuzzy-set based method, and Bayesian-based method
optimize.kmer()	Optimize kmer length by trying kmers which length is in the range from 1 to max.kmer. The optimal kmer will have maximal species identification success rate
consensus.identify	Make consensus for identifications from two or more methods, usually for a set of query sequences
compare2delimitations()	Comparison between two delimitations of a group of samples, for instance, traditionally morphological delimitation and molecular delimitation (MOTU)
save.ids()	Results or output of species identification function is saved in a file using this function

References

Brown SDJ, Collins RA, Boyer S, Lefort M-C, Malumbres-Olarte J, Vink CJ, Cruickshank RH (2012) Spider: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. Mol Ecol Resour 12:562–565

Charif D, Lobry JR (2007) SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In: Structural approaches to sequence evolution: molecules, networks, populations. Springer Verlag, New York, pp 207–232

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376

Jombart T (2008) Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24(11):1403–1405

Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290

R Development Core Team (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna https://cran.r-project.org/

Ratnasingham S, Hebert PDN (2007) BOLD: the barcode of life data system. Mol Ecol Notes 355–364. www.barcodinglife.org

Saitou N, Nei M (1987) The neighbor-joining method—a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Tamura K, Nei M (1993) Estimation of the number of nucleotide sub-stitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10(3):512–526

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Zhang A-b, Hao M-d, Yang C-q, Shi Z-y (2016) BarcodingR: an integrated r package for species identification using DNA barcodes. Methods Ecol Evol. https://doi.org/10.1111/2041-210X.12682

Implications and Utility of DNA Barcoding

J. Suriya, M. Krishnan, S. Bharathiraja, V. Sekar, and V Sachithanandam

Abstract Classical way of practicing taxonomy is in endangered race in the era of genomics. In recent years taxonomy became fashionable that is owing to the revolutionary approaches in taxonomy called DNA barcoding. It is a novel approach that has generated optimism in enhancing biodiversity assessments. In DNA barcoding, complete data can be retrieved from a single specimen regardless of life stage or morphological characters. The core idea behind the DNA barcoding is the fact of minor variation in highly conserved region of DNA during the evolution within the species. Sequences have successfully been utilized for DNA barcoding which include cytoplasmic mitochondrial DNA (cox1), chloroplast DNA (trnL-F, matK, ndhF, atpB and rbcL) and nuclear DNA (ITS and housekeeping genes). Now it has been used for diverse applications such as biodiversity assessment, life history and ecological studies, forensic analysis and many more. In this chapter, we discuss the significance and utility of DNA barcoding in various fields.

Keywords DNA barcoding · Molecular taxonomy · Biodiversity · Mitochondrial genes · Gene sequencing

J. Suriya (✉) · M. Krishnan
Department of Environmental Biotechnology, School of Environmental Sciences, Bharathidasan University, Tiruchirappalli, Tamil Nadu, India

S. Bharathiraja
Department of Biomedical Engineering, Nano-Biomedicine Lab, Pukyong National University, Nam-Gu, Busan, South Korea

V. Sekar · V. Sachithanandam
National Centre for Sustainable Coastal Management, Ministry of Environment, Forests & Climate Change, Chennai, Tamil Nadu, India

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_3

1 Introduction

Earth comprises of innumerable diversity of organisms which are mostly uncharacterized (Wilson 2003). The classical techniques used for species identification pave way for many challenges and problems; henceforth, the taxonomist became an endangered race in the era of genomics. Molecular techniques are promising approaches for species identification because it overcomes the difficulties in the morphology-based identification methods. Now taxonomy became fashionable owing to the innovative approaches in taxonomy called DNA barcoding. DNA barcodes usually comprise of short sequences of DNA that can be amplified with the aid of polymerase chain reaction (PCR) and then sequenced for analysing the species of interest. It has been successfully utilized for the identification of subspecies, eco- and morpho-types, cultivars, mutants, clones and species complex. Molecular operational taxonomic units (MOTU) is used for the identification of dissimilar species, whereas similar species are identified by comparing the barcode of a species of interest with sequences retrieved from other taxa (Floyd et al. 2002). It is used for barcoding of various organisms including freshwater meiobenthos (Markmann and Tautz 2005), soil meiofauna (Blaxter et al. 2004), extinct birds (Lambert et al. 2005), marine organisms (Shander and Willassen 2005), fishes (Ward et al. 2005) and insects (Smith et al. 2005). In particular, DNA barcodes are more potentially applied in biosecurity such as law enforcement and primatology (Lorenz et al. 2005) as well as for surveillance of disease vectors (Besansky et al. 2003) and invasive insects (Armstrong and Ball 2005). Among other species identification approaches, DNA barcoding is a promising technique for species identification due to its expert-authenticated verification system and accuracy (Ajmal Ali et al. 2014).

The two main aim of DNA barcoding are (1) discovery of new species and its identification (2) assign of unknown specimens to species (Hebert et al. 2003). Figure 1 illustrates the different steps involved in the DNA barcoding approach.

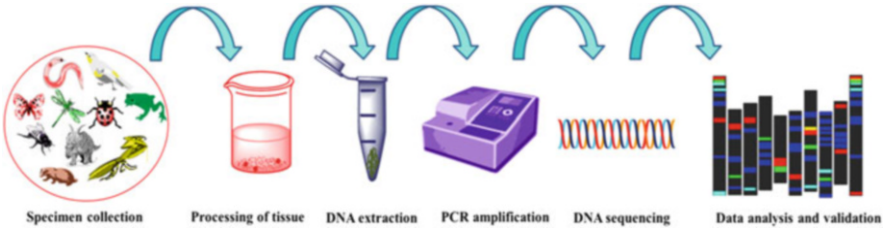


Fig. 1 Steps involved in DNA barcoding process—from specimen to barcode data analysis

2 Significance of DNA Barcoding

Specimen identification using barcoding is usually attained by obtaining short DNA sequence—the 'barcode'—from the genome of the species. Then the barcode sequence of the unknown species is compared with the reference library of barcode sequences of known species. If the sequence of the unknown species is closely matches with the sequences in the reference library, a species is identified. Novel barcode sequence implies the findings of new species. The DNA barcode gives great support to innumerable scientific domains such as epidemiology, ecology, biomedicine, evolutionary biology, biogeography and conservation biology and also in bioindustry. For handling large sampling groups, DNA barcoding is the most preferred technique due to its time and cost effectiveness. In addition to this, DNA barcoding is also employed in the isolation and identification of unknown species which have significance in the medical, ecological and agronomical fields (Armstrong and Ball 2005; Ball and Armstrong 2006). Moreover, it is used for the detection of patented organisms in agro-biotechnology, either to protect intellectual property rights for bioresources (Taberlet et al. 2007) or certify the source organism (Rastogi et al. 2007).

The main advantage of DNA barcoding is the acquiring large volume of molecular data. It is particularly useful in determining the identity of damaged organisms. The DNA barcoding is potentially utilized in the diet analyses, food industry and forensic sciences in preventing poaching of endangered species and illegal trade. Shark fin trade is the best example for illegal trade in several coastal African countries which is quite thread to the biodiversity conservation. Moreover, DNA barcoding also play a crucial role in environmental law enforcement.

3 Gene Markers in DNA Barcoding

Different gene regions such as mitochondrial DNA (*COI*), nuclear DNA (*ITS*) and chloroplast DNA (*rbcL*, *trnL-F*, *matK*, *psbA*, *trnH* and *psbK*) have been used for DNA barcoding approach; however, cytochrome *c* oxidase is a universal DNA barcode for animals (Ajmal Ali et al. 2014). More than 95% of species have unique COI barcode sequences. The COI gene has been successfully utilized for the identification of abundance of species such as birds (Kerr et al. 2007), butterflies (Hajibabaei et al. 2006), gastropods (Remigio and Hebert 2003), ants (Smith et al. 2005), springtails (Hogg and Hebert 2004), Protista (Evans et al. 2007), spiders (Greenstone et al. 2005), Crustacea (Costa et al. 2007) and fish (Ward et al. 2005). In plants, mitochondrial DNA revealed gradual substitution rates than animals and also showed intramolecular recombination. Henceforth, *rpoB*, *rpoC1 272* and *rbcL* or a section of *matK* markers are used for plant identification owing to its higher rate of evolution.

4 Applications of DNA Barcoding

4.1 DNA Barcoding in Biodiversity Assessments and Conservations

Barcoding helps to identify the species quickly and cheaply. According to Hebert et al. (2010) estimation, 0.1% of all described animal species are sequenced within 90 min at a cost of $2–5 per specimen, whereas morphological analyses required several months of field work at a cost of $100 per species (Hebert and Gregory 2005; Stoeckle and Hebert 2008). Genetic tools are more useful in the identification of cryptic and invasive species (Stoeckle 2008) and management of coral reefs (Neigel et al. 2007) and fisheries (Swartz et al. 2008). Discovery of cryptic diversity play an important role in new species identification by DNA barcoding. DNA barcoding is also used for the identification of new species (Pauls et al. 2010). It identifies the innumerable species that remain unexplored (Wilson 1994). For adequate description of the species, DNA barcoding along with traditional taxonomic methods is used (Prendini 2005).

DNA barcoding conserves the species by enhancing local biodiversity assessments to give priority for conservation areas, by giving data about phylogenetic diversity and evolutionary histories. The main advantage of DNA barcoding in biodiversity conservation is assessing the biodiversity in a time- and cost-effective manner where financial resources are limited. This is particularly crucial because the majority of described biodiversity is in developing countries, where resources for biodiversity assessments are lacking. Francis et al. (2010) and Ward et al. (2008) have conserved several species with the aid of barcoding in Southeast Asia and Australia, respectively, and also Neigel et al. (2007) identified many juvenile and larval species for which morphological data is absent.

4.1.1 DNA Barcoding on Bacteria

Most of the bacterial species are cryptic in nature; henceforth DNA barcoding holds promise to identify bacteria at species level (Begerow et al. 2010). Phytoplasmas are bacterial phytopathogens which cause losses in agricultural production. Makarova et al. (2012), used elongation factor Tu (tuf) gene to develop a universal DNA barcode for phytoplasma identification. Portion of the DnaA replication initiation factor (RIF) has been used to discriminate plant-associated bacterial pathogens at the species level in *Xanthomonas*, *Pseudomonas* and *Xylella* (Schneider et al. 2011). Natalie (2013) used DNA barcoding to evaluate the consequences of fire on microbial communities in chaparral soils and also compared the microbes in the burnt and unburned soil samples. Thao et al. (2013) used fluorescent proteins to differentiate selected enterobacterial species rapidly in DNA barcoding. Multilocus sequence typing (MLST) approaches are widely applied for prokaryote identification like *Wolbachia* strains (Baldo et al. 2006). DNA barcode approaches produce phylogenetic hypotheses for

entire bacterial communities with the aid of universal primers (Wang et al. 2012a) which allow the ecologists to give answer for the fundamental questions regarding their distribution and assembly (Mouquet et al. 2012). Links et al. (2012) evaluated the efficiency of 16S rRNA and cpn60 genes as bacterial barcodes. They suggested that the cpn60 is the suitable barcode for bacteria than 16S rRNA due to its large barcode gap. However, 16S rRNA gene is mostly used as the barcode for the identification of bacterial and archaeal communities. Enormous studies have been used this molecular marker as a barcode to quantify bacterial community from environmental samples (Hugenholtz et al. 1998). Those results implied that we have only studied little part of the bacterial community that exist in the environments.

4.1.2 DNA Barcoding on Fungi

Identification of fungi by morphological characters needs well-trained experts due its very complex nature (filamentous fungi). Fungi occupy one of the largest kingdoms; however, large numbers of them are not described yet (Mora et al. 2011). It is very tricky to identify numerous fungi due its unculturable nature (Begerow et al. 2010). In the clinical laboratory, proper identification of fungal pathogens continues to be a difficult task. It is crucial to identify disease-causing agents for detecting novel therapeutic agents. In such cases molecular identification holds promise in the identification of fungal species. Various genetic loci mostly used for fungi identification include 28S rRNA large subunit (LSU) (Scorzetti et al. 2002), β-tubulin (Balajee et al. 2007), internal transcribed spacer (ITS) of the ribosomal DNA (rDNA) (Rainer and de Hoog 2006) and multilocus sequencing (Ngamskulrungroj et al. 2009). Various essential fungal groups were identified using different methods on the basis of their sequence variation (Cerqueira et al. 2014). Several studies indicated that COI is suitable only for few genera such as *Penicillium* (Seifert et al. 2007). Multiple copies of COI gene are present in the *Fusarium*, *Aspergillus niger* and *Basidiomycota* groups; henceforth COI is unfit for identification of fungi at the species level (Gilmore et al. 2009; Vialle et al. 2009). In addition, PCR amplification and sequencing are very difficult (Dentinger et al. 2011) and needed nested primers for amplification of the entire region of COI (Gilmore et al. 2009). These results pave the way for the use of multiple genetic loci for fungal species identification (Roe et al. 2010). Multigene approaches are commonly used for phylogenetic studies. Geiser et al. (2007) used diverse sequences such as ITS, cox1, calmodulin and β-tubulin for black aspergilli species identification, and they reported that either calmodulin or β-tubulin could be a suitable barcode for species identification.

4.1.3 DNA Barcoding on Terrestrial Species

Among terrestrial ecosystems, soil represents one of the most diversified habitats with majority of species that form terrestrial biodiversity. About 25% of the organisms on earth is represented by soil organisms (Decaëns et al. 2013). One gramme of soil may

host several thousands of diverse bacterial and fungal species, and a square metre of soil may contain several hundreds of different arthropods species (Decaëns et al. 2013). Floyd et al. (2002) employed DNA barcoding approach for identification of soil nematodes. Hebert et al. (2004) identified 40% North American avian species with the aid of barcoding. Following this study, several scientist employed DNA barcoding tool to assess avian species of Neotropics (Kerr et al. 2009), the Palearctic (Johnsen et al. 2010), the Indomalayan region (Lohman et al. 2010), and Australasia (Patel et al. 2010). Canadian insects were described by Hebert et al. (2016) through barcoding approach. Francis et al. (2010) evaluated the potential of DNA barcodes to understand and to conserve mammals' diversity in Southeast Asia, and they found that within the region 50% of mammal species richness is underestimated. Moreover, this approach has been shown to be effective in the identification of earthworms (Chang and James 2011). All of them described a highly structured variation of the barcode in amphibians (Smith et al. 2008a), reptiles (Nagy et al. 2012), small mammals (Lu et al. 2012), bats (De Pasquale and Galimberti 2014), squirrels (Ermakov et al. 2015), primates (Ruiz-García et al. 2014) and rodents (Galan et al. 2012).

4.1.4 DNA Barcoding on Freshwater Species

Fresh water occupies only a relatively small proportion of the earth; however, freshwater habitats are home to innumerable species (Dudgeon et al. 2006) and extensively used by human beings. Henceforth, it is very crucial to recurrently evaluate their ecological health and develop potential monitoring approaches. Rossi and Mantelatto (2013), Udayasuriyan et al. (2015) and Jose et al. (2016) used MT-COI gene for identification of freshwater prawns. Finn et al. (2014) discovered the loss of a baetid mayfly due to the loss of alpine glaciers. Hurwood et al. (2014) applied mtDNA sequences to explore the freshwater prawn diversity from western India to western Java. Lim et al. (2016) and Kermarrec et al. (2014) utilized NGS to assess metazoan and diatom communities, respectively. A large number of scientists such as Hubert et al. (2008), Nwani et al. (2011), Bhattacharjee and Ghosh (2014), Chakraborty and Ghosh (2014), Lakra et al. (2016) and Keskin and Atar (2013) applied DNA barcoding for identification of freshwater fishes, whereas Corse et al. (2010), Carreon-Martinez et al. (2011) and Jo et al. (2014) employed DNA barcoding to find out freshwater carnivorous fish diets.

4.1.5 DNA Barcoding of Marine Species

Marine has rich biodiversity compared to terrestrial and freshwater ecosystem, because it covers more than 70% of our planet. The marine ecosystem is the suitable habitat for many micro- and macro-organisms, both flora and fauna. Thirty four animal phyla are found in the marine, among the total of 35 animal phyla (Gray 1997). The occurrence of cryptic species is prevalent in marine ecosystems. The major difficulty in the marine ecosystem is the connecting larval stages with the adult

forms. DNA barcoding can effectively connect the larval stages to adult forms. DNA barcoding has been utilized to find out invasive species threat to marine biodiversity (Molnar et al. 2008). Indicator species barcoding can be successful in the assessment and reduction of marine pollution including coastal pollution. The main aim of barcoding is the discovery of new species.

Seagrasses possess several valuable secondary compounds like phenolic, rosmarinic and zosteric acids which are used in traditional medicines as antioxidants and potential antifouling agent. Various molecular markers are used for seagrass identification such as trnK introns and rbcL for *Zostera* (Les et al. 2002), nuclear ITS for *Halophila* (Waycott et al. 2009) and ITS1, 5.8S rDNA and ITS2 for *Halophila* (Uchimura et al. 2008). Mangroves have enormous economic and ecological significance. This unique and dynamic habitat is increasingly depleted and threatened by natural as well as man-made disaster. So, we are in urge to conserve this ecosystem. DNA barcoding serves to produces phylogenetic information which is helpful in developing unified mangrove management plan worldwide (Daru et al. 2013). According to UNESCO world heritage list, Sunderbans is the only largest halophytic mangrove forest (http://whc.unesco.org/en/list) in the world. Large number of estuarine and marine species comes to this ecosystem to breed, and the juveniles stay back to utilize its natural resources (Trivedi et al. 2013).

Morphological techniques are ineffective in the identification of different species of red marine macro algae. Usually mitochondrial COI gene and 23S rRNA's UPA (universal plastid amplicon) domain V gene are used for detection of diverse species of Kallymeniaceae family red alga. COI is a more sensitive marker and pave the way for discovery of *Euthora timburtonii* (Clarkston and Saunders 2010). Moreover, it is very tricky to identify *Gracilaria* morphologically, and DNA barcoding can be very effective to identify *Gracilaria* in species level (Kim et al. 2010). Genes of 16S rRNA and 23S rRNA were used as the barcode for isolation and identification of a novel microalga from Indian Ocean (Ahmad et al. 2013).

The pteropods are ecologically important marine species due to its susceptibility to ocean acidification. *Diacavolinia* pteropods' barcoding showed that the Atlantic specimens contain only single monophyletic species and also revealed species-level difference between Pacific and Atlantic populations (Maas et al. 2013). Two hundred and twenty seven Canadian marine mollusc species were identified with the aid of DNA barcoding (Layton et al. 2014). Chen et al. (2011) potentially utilized barcoding approach to identify 60 venerid species from the coast of China. Marine oyster has significant economic value. Oysters are mostly differentiated on the basis of their shell morphology; however, the habitat alters the shell morphology greatly (Tack et al. 1992). In such cases, DNA barcoding holds promise to identify oyster at species level. Sponge Barcoding Project, http://www.spongebarcoding.org, is much helpful for developing workflow to scrutinize large number of sponge.

FISH-BOL (http://www.fishbol.org) and SharkBOL (http://www.sharkbol.org) are the two main fish global barcoding initiatives. DNA barcoding is not only efficient for the identification of whole fish, but it also very helpful in the identification of eggs, larvae, fins and fillets which are very difficult to identify based on their morphology. Three hundred and ninety one ornamental fish species isolated from 8 coral reef

locations were identified with the help of DNA barcoding (Steinke et al. 2009). Many scientists have potentially utilized DNA barcoding approach for the identification of marine fish throughout the world (Trivedi et al. 2014; Ardura et al. 2013; Weigt et al. 2012). Mammalia Barcode of Life (http://www.mammaliabol.org) is launched for barcoding of mammals including the marine mammals. Alfonsi et al. (2013) carried out the study along the French Atlantic coast and found that DNA barcoding combined with a stranding network can be very useful in monitoring marine mammal diversity.

4.2 Metabarcoding

In recent years soil metabarcoding draws the attention of the scientist and results in large number of publications. This is because of easy handling and reduction of cost and time in NGS platforms, and also many bioinformatics pipelines are developed for analysis of metadata (Yang et al. 2013). Researchers have successfully utilized this approach to shed light on innumerable hitherto unexplored biodiversity from unexplored areas. The two main objectives of the soil biodiversity are (1) to assess the soil biodiversity's structure and functions, for example, ecological roles and (2) to obtain the vacillations of soil biodiversity in different environmental conditions for ascertaining protective measurements. These two objectives can be easily fulfilled by using metaborcoding.

For example metabarcoding is used for assessing the spatial distribution of soil biodiversity. The soil DNA can be stored for prolonged periods of time (Lauber et al. 2010) which offers to obtain the soil biodiversity over time in relation to climatic change (Dumbrell et al. 2011).

4.3 Detection of Mislabelling and Food Piracy

DNA barcoding is potentially applied for identification, authentication and safety assessment of food, particularly for cooked, processed or smoked products. This molecular approach allows us to detect the origin of certain food products (Galimberti et al. 2013). Lowenstein et al. (2009) collected Japanese delicacy tuna sushi from various US restaurants and find out the presence of endangered species, fraud and also a health hazard. Two hundred and fifty four Canadian seafood samples were analysed for mislabelling, and the results divulged that 41% of the samples were mislabelled (Hanner et al. 2011). Holmes et al. (2009) identified shark fins that were illegally marketed by fishers in Australia. A study was conducted on 62 seafood samples in Malaysia to ascertain mislabelling. Among these samples, 16% were mislabelled (Chin et al. 2016). Wong and Hanner (2008) detected the mislabelling of 25% of the North American seafoods with the aid of DNA barcoding.

DNA barcodes have been used to identify the mislabelling of toxic puffer fish in a Chicago market as monkfish (Cohen et al. 2009). It has been used by several

researchers as a forensic tool for the traceability of edible fish (Barbuto et al. 2010; Smith et al. 2008b). Newmaster et al. (2013) investigated herbal product substitution and contamination to protect the consumers from health risks associated with food adulteration. Among 44 samples, 30 samples were substituted, and only two companies had herbal products without any substitution among 12 companies. Herbal tea contamination was detected by Stoeckle et al. (2011) using DNA barcoding. Parvathy et al. (2014, 2015) have potentially used DNA barcoding for detecting chilli and plant-based adulteration in black pepper powder and in turmeric powder. DNA barcoding-based species identification is applied to the verification/certification of mushroom-containing dietary supplements (Raja et al. 2017). Vassou et al. (2015) collect 13 species of *Sida* from market samples for identification of *Sida cordifolia*, used for treating neurological disorders. They found that none of the market samples belonged to the *S. cordifolia.* These types of substitutions not only fail to give the expected effect but may also give undesirable health effects. COX1 is employed for mammalian meat traceability (Luo et al. 2011) and avian meat product identification (Hebert et al. 2004).

4.4 Controlling Agricultural Pest and Predators

DNA barcoding is much helpful for agricultural scientists in the exact and rapid identification of agriculturally important insect pests and their predators. DNA barcoding was used by Li et al. (2012) to identify Noctuidae pests. Smith et al. (2013) reported that more than 20,000 sequences of microgastrine wasps were produced by scientists. Nagoshi et al. (2011) identified armyworms in Florida with the aid of DNA barcoding. This approach is also applied for identification of scale insects (Kondo et al. 2008), *Ectoedemia* (Nieukerken et al. 2012), Nearctic Muscidae (Renaud et al. 2012), storage pests (Yang et al. 2013) and aphids' cryptic species differentiation (Rebijith et al. 2013). Mostly spiders are used to regulate the population of insect pests. Barrett and Hebert (2005) used DNA barcoding for spider identification. Marie-Stephane et al. (2012) utilized COI, Cytb, 12SrRNA and ITSS markers for *Typhlodromus pyri* populations' identification. European beetle in particular German fauna was assigned to known species by using CO1 gene (Hendrich et al. 2015). Woodcock et al. (2013) assessed the beetle diversity in Canada, Manitoba and Churchill to develop barcode library for subarctic region beetles by using DNA molecular markers for management of pest control.

4.5 Identifying Disease Vectors and Parasites

Accurate identification of parasites and their vectors is crucial in vector-borne disease surveillance programmes. It is very difficult to identify many parasites morphologically; in addition, morphological expertise in parasite and its vector

identification is scarce (Besansky et al. 2003). DNA barcoding overwhelm this problem which allows non-taxonomists to identify disease vectors and to understand and control disease-carrying pathogens and pests (http://www.barcoding.si.edu/PDF/CBOL). Disease-causing vectors such as mosquitoes are potentially identified with the help of DNA barcoding (Wang et al. 2012b; Kumar et al. 2013; Ashfaq et al. 2014). Kumar et al. (2007) confirmed 62 mosquito species in DNA barcoding among 63 morphologically different species from India. Bourke et al. (2013) reported that DNA barcoding is a potential tool in the identification of diverse species of Brazilian mosquitoes.

Dhananjeyan et al. (2010) and Naddaf et al. (2012) also successfully employed DNA-based method molecular markers for the identification of mosquitoes in India and Iran, respectively. CO1 gene was used for the identification of damaged specimen of German mosquitoes with reference to three *Aedes* species reported in Europe (Kruger et al. 2014).

4.6 Identification of Medical Plants

Medicinal plants were used as preventive/protective agent for various diseases; people use medicinal plants throughout the world. However, adulteration is the major problem for users, and it's important to authenticate the plat species. DNA barcoding is a promising tool in the authentification of medicinal plants to discriminate the good one from adulterants. Vassou et al. (2015) applied DNA barcoding for identification of *Sida cordifolia. Peucedanum praeruptorum*, a traditional medicinal plant, was identified by Zhou et al. (2014) using DNA barcoding. *Ginkgo biloba*, exploited for treating dementia, Alzheimer's disease and Parkinson's disease, was authenticated by using DNA barcoding (Little 2014). An antimutagenic and antioxidant fruit *Phoenix dactylifera was* distinguished by Enan and Ahamed (2014). Hou et al. (2013) applied DNA barcoding for identification of traditional Chinese medicinal plant *Lonicera japonica*. Wong et al. (2013) evaluated the efficiency of seven DNA barcodes for discriminating closely related medicinal *Gentiana* species and their adulterants. MATK gene was used as a molecular marker for characterization of *Croton bonplandianum* (Chandramohan et al. 2013), whereas Rai et al. (2012) used *ITS2* gene for detecting substitution in the medicinal plant *Asparagus racemosus* and for identification of *Cinnamomum osmophloeum*, respectively.

4.7 Monitoring Water Quality

Diatoms are commonly used for the assessments of water quality; however, species level identification is very difficult because it needs in-depth knowledge of the organisms under investigation, and it is also a time-consuming approach. To overcome this bottleneck, Zimmermann et al. (2015) applied metabarcoding of diatoms

via NGS sequencing for evaluation of water quality. DNA barcodes of stream macroinvertebrates were used for monitoring water quality by Sweeney et al. (2011). Macroinvertebrate diversity is used as indicator for assessing ecosystem health in AZTI's Marine Biotic Index (AMBI). Genetic-based AMBI was used for faster and cheaper marine monitoring and health assessment of marine ecology (Aylagas et al. 2014).

4.8 DNA Barcoding on Forensic Sciences and Conservation of Endangered Species

Application of DNA barcoding in the investigations of animal cruelty and poaching and illegal collection and trade of flora and fauna has been increased recently. Naro-Maciel et al. (2010) employed DNA barcoding for globally threatened marine turtles. DNA barcoding is a potential tool for detecting poaching of Indian peafowl, Chinese sika deer subspecies, roe deer, guanaco, crocodile, reedbuck, lowland tapir and Cypriot mouflon (Gupta et al. 2005; Wu et al. 2005; An et al. 2007; Marín et al. 2009; Eaton et al. 2010; Dalton and Kotze 2011; Sanches et al. 2011; Barbanera et al. 2012). Indo-Chinese spitting cobra, scarlet macaw, sturgeons and paddlefish, elephant and Southeast Asian monitor lizards were detected by several researchers with the aid of DNA barcoding (Shivji et al. 2005; Baker et al. 2010; Filonzi et al. 2010; Gupta et al. 2011). Gupta et al. (2011) also employed DNA barcoding for detecting illegal killing of tiger.

DNA barcoding can be used to identify endangered sea turtles by assessing turtle meat, carcasses or eggs that are illegally traded (Vargas et al. 2009). Laramie et al. (2015) used DNA barcoding for evaluating the distribution of an endangered salmonid in Methow and Okanogan Subbasins of the Upper Columbia River, which span the border between Washington, USA, and British Columbia, Canada. Eva et al. (2016) detected endangered Mekong giant catfish *Pangasianodon gigas* by applying DNA barcode technique. Panprommin and Panprommin (2017) assessed molecular markers for identification of a critically endangered species like *Trigonostigma somphongsi* in Thailand. Bhattacharyya et al. (2015) employed ISSR and DAMD molecular markers for assessing *Dendrobium nobile*, an endangered medicinal orchid. COI was utilized by Dubey et al. (2011) for identification of some endangered Indian snake species.

5 Conclusion

DNA barcoding is mainly focused on animals, and it is high time for giving importance to barcoding of plants and protists. The major drawback of plant barcoding is the absence of universal barcode gene in plants. In spite of some bottlenecks, DNA barcoding approach has successfully been applied for assessing and conserving

biodiversity in the massive and diverse ecosystem, detecting mislabelling and food piracy, agricultural pest controlling, disease vectors identification, medical plants identification, conservation of endangered species and monitoring water quality and also employed in forensic palynology. In conclusion, DNA barcoding is a crucial approach which expands our knowledge by assessing innumerable species in inexpensive and time-effective manner. It increases the communication between diverse scientific communities such as, phylogeneticists, taxonomists and population geneticists.

Acknowledgements The authors are grateful to the University Grants Commission (UGC), New Delhi—Dr. D.S. Kothari Post-Doctoral Fellowship—for their financial support.

References

Ahmad I, Fatma Z, Yazdani SS (2013) DNA barcode and lipid analysis of new marine algae potential for biofuel. Algal Res 2:10–15

Ajmal Ali M, Gyulai G, Hidvégi N, Kerti B, Al Hemaid FM, Pandey AK, Lee J (2014) The changing epitome of species identification – DNA barcoding. Saudi J Biol Sci 21:204–231

Alfonsi E, Meheust E, Fuchs S (2013) The use of DNA barcoding to monitor the marine mammal biodiversity along the French Atlantic coast. In: Nagy ZT, Backeljau T, De Meyer M, Jordaens K (eds) DNA barcoding: a practical tool for fundamental and applied biodiversity research, vol 365. ZooKeys, pp 5–24

An J, Lee M-Y, Min M-S, Lee M-H, Lee H (2007) A molecular genetic approach for species identification of mammals and sex determination of birds in a forensic case of poaching from South Korea. Forensic Sci Int 167:59–61

Ardura A, Planes S, Garcia-Vazquez E (2013) Applications of DNA barcoding to fish landings: authentication and diversity assessment. In: Nagy ZT, Backeljau T, De Meyer M, Jordaens K (eds) DNA barcoding: a practical tool for fundamental and applied biodiversity research, vol 365. ZooKeys, pp 49–65

Armstrong KF, Ball SL (2005) DNA barcodes for biosecurity: invasive species identification. Philos Trans R Soc Lond Ser B Biol Sci 360(1462):1813–1823

Ashfaq M, Hebert PDN, Mirza JH, Khan AM, Zafar Y, Mirza S (2014) Analyzing mosquito (Diptera: Culicidae) diversity in Pakistan by DNA barcoding. PLoS One 9:e97268

Aylagas E, Borja A, Rodríguez-Ezpeleta N (2014) Environmental status assessment using DNA metabarcoding: towards a genetics based marine biotic index (gAMBI). PLoS One 9(3):e90529

Baker CS, Steel D, Choi Y, Lee H, Kim KS, Choi SK et al (2010) Genetic evidence of illegal trade in protected whales links Japan with the US and South Korea. Biol Lett 6(5):647–650. https://doi.org/10.1098/rsbl.2010.0239

Balajee SA, Houbraken J, Verweij PE, Hong SB, Yaghuchi T, Varga J, Samson RA (2007) *Aspergillus* species identification in the clinical setting. Stud Mycol 59:39–46

Baldo L, Hotopp JCD, Jolley KA et al (2006) Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. Appl Environ Microbiol 72(11):7098–7110

Ball SL, Armstrong KF (2006) DNA barcodes for insect pest identification: a test case with tussock moths (Lepidoptera: Lymantriidae). Can J For Res 36:337–350

Barbanera F, Guerrini M, Beccani C, Forcina G, Anayiotos P, Panayides P (2012) Case report. Conservation of endemic and threatened wildlife: molecular forensic DNA against poaching of the Cypriot mouflon (Ovis orientalis ophion, Bovidae). Forensic Sci Int Genet 6:671–675

Barbuto M, Galimberti A, Ferri E, Labra M, Malandra R, Galli P et al (2010) DNA barcoding reveals fraudulent substitutions in shark seafood products: the Italian case of "palombo" (Mustelus spp.). Food Res Int 43:376–381

Barrett RDH, Hebert PDN (2005) Identifying spiders through DNA barcodes. Can J Zool 83:481–491

Begerow D, Nilsson H, Unterseher M, Maier W (2010) Current state and perspectives of fungal DNA barcoding and rapid identification procedures. Appl Microbiol Biotechnol 87(1):99–108

Besansky NJ, Severson DW, Ferdig MT (2003) DNA barcoding of parasites and invertebrate disease vectors: what you don't know can hurt you. Trends Parasitol 19:545–546

Bhattacharjee MJ, Ghosh SK (2014) Design of mini-barcode for catfishes for assessment of archival biodiversity. Mol Ecol Resour 14:469–477

Bhattacharyya P, Kumaria S, Tandon P (2015) Applicability of ISSR and DAMD markers for phyto-molecular characterization and association with some important biochemical traits of Dendrobium nobile, an endangered medicinal orchid. Phytochemistry 117:306–316

Blaxter M, Elsworth B, Daub J (2004) DNA taxonomy of a neglected animal phylum: an unexpected diversity of tardigrades. Proc R Soc B 271:189–192

Bourke BP, Oliveira TP, Suesdek L, Bergo ES, Sallum MA (2013) A multi-locus approach to barcoding in the Anopheles strodei subgroup (Diptera: Culicidae). Parasit Vectors 6:111

Carreon-Martinez L, Johnson TB, Ludsin SA, Heath DD (2011) Utilization of stomach content DNA to determine diet diversity in piscivorous fishes. J Fish Biol 78:1170–1182

Cerqueira GC, Arnaud MB, Inglis DO, Skrzypek MS, Binkley G, Simison M, Miyasato SR, Binkley J, Orvis J, Shah P, Wymore F, Sherlock G, Wortman JR (2014) The *Aspergillus* genome database: multispecies curation and incorporation of RNA-Seq data to improve structural gene annotations. Nucleic Acids Res 42:D705–D710

Chakraborty M, Ghosh SK (2014) An assessment of the DNA barcodes of Indian freshwater fishes. Gene 537:20–28

Chandramohan A, Divya SR, Dhanarajan MS (2013) Matk gene based molecular characterization of medicinal plant—Croton bonplandianum Baill. Int J Biosci Res 2:1–7

Chang C-H, James S (2011) A critique of earthworm molecular phylogenetics. Pedobiologia 54S: S3–S9

Chen J, Li Q, Kong L (2011) How DNA barcodes complement taxonomy and explore species diversity: the case study of a poorly understood marine fauna. PLoS One 6(6):e21326

Chin TC, Bakar AA, Abidin DHZ, Nor SAM (2016) Detection of mislabelled seafood products in Malaysia by DNA barcoding: improving transparency in food market. Food Control 64:247–256

Clarkston BE, Saunders GW (2010) A comparison of two DNA barcode markers for species discrimination in the red algal family Kallymeniaceae (Gigartinales, Florideophyceae), with a description of Euthora timburtonii sp. Botany 88:119–131

Cohen NJ, Deeds JR, Wong ES, Hanner RH, Yancy HF, White KD et al (2009) Public health response to puffer fish (Tetrodotoxin) poisoning from mislabeled product. J Food Prot 72:810–817

Corse E, Costedoat C, Chappaz R, Pech N, Martin J, Gilles A (2010) A PCR-based method for diet analysis in freshwater organisms using 18S rDNA barcoding on faeces. Mol Ecol Resour 10:96–108

Costa FO, de Waard JR, Boutillier J, Ratnasingham S, Dooh RT, Hajibabaei M, Hebert PDN (2007) Biological identifications through DNA barcodes: the case of the Crustacea. Can J Fish Aquat Sci 64:272–295

Dalton DL, Kotze A (2011) DNA barcoding as a tool for species identification in three forensic wildlife cases in South Africa. Forensic Sci Int 207:e51–e54

Daru BH, Yessoufou K, Mankga LT, Davies TJ (2013) A global trend towards the loss of evolutionarily unique species in mangrove ecosystems. PLoS One 8(6):e66686

De Pasquale PP, Galimberti A (2014) New records of the Alcathoe bat, *Myotis alcathoe* (Vespertilionidae) for Italy. Barbastella 7:1

Decaëns T, Porco D, Rougerie R, Brown GG, James SW (2013) Potential of DNA barcoding for earthworm research in taxonomy and ecology. Appl Soil Ecol 65:35–42

Dentinger BTM, Didukh MY, Moncalvo J-M (2011) Comparing COI and ITS as DNA barcode markers for mushrooms and allies (*Agaricomycotina*). PLoS One 6:e25081

Dhananjeyan KJ, Paramasivan R, Tewari SC, Rajendran R, Thenmozhi V, Jerald V, Leo S et al (2010) Molecular identification of mosquito vectors using genomic DNA isolated from egg-shells, larval and pupal exuvium. Trop Biomed 27:47–53

Dubey B, Meganathan PR, Haque I (2011) DNA mini-barcoding: an approach for forensic identification of some endangered Indian snake species. Forensic Sci Int Genet 5:181–184

Dudgeon D, Arthigton AH, Gessner MO et al (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. Biol Rev Camb Philos Soc 81:163–182

Dumbrell AJ, Ashton PD, Aziz N, Feng G, Nelson M, Dytham C, Fitter AH, Helgason T (2011) Distinct seasonal assemblages of arbuscular mycorrhizal fungi revealed by massively parallel pyrosequencing. New Phytol 190:794–804

Eaton MJ, Meyers GL, Kolokotronis S-O, Leslie MS, Martin AP, Amato G (2010) Barcoding bushmeat: molecular identification of Central African and South American harvested vertebrates. Conserv Genet 11:1389–1404

Enan MR, Ahamed A (2014) DNA barcoding based on plastid matK and RNA polymerase for assessing the genetic identity of date (Phoenix dactylifera L.) cultivars. Genet Mol Res 13:3527–3536

Ermakov OA, Simonov E, Surin VL, Titov SV, Brandler OV, Ivanova NV, Borisenko AV (2015) Implications of hybridization, NUMTs, and overlooked diversity for DNA barcoding of Eurasian ground squirrels. PLoS One 10(1):e0117201

Eva B, Harmony P, Thomas G, Francois G, Alice V, Claude M, Tonya D (2016) Trails of river monsters: detecting critically endangered Mekong giant catfish Pangasianodon gigas using environmental DNA. Glob Ecol Conserv 7:148–156

Evans KM, Wortley AH, Mann DG (2007) An assessment of potential diatom "barcode" genes (cox1, rbcL, 18S and ITS rDNA) and their effectiveness in determining relationships in Sellaphora (Bacillariophyta). Protist 158:349–364

Filonzi L, Chiesa S, Vaghi M, Marzano FN (2010) Molecular barcoding reveals mislabeling of commercial fish products in Italy. Food Res Int 43:1383–1388

Finn DS, Zamora-Muñoz C, Múrria C, Sáinz-Bariáin M, Alba-Tercedor J (2014) Evidence from recently deglaciated mountain ranges that *Baetis alpinus* (Ephemeroptera) could lose significant genetic diversity as alpine glaciers disappear. Freshw Sci 33(1):207–216

Floyd R, Abebe E, Papert A, Blaxter M (2002) Molecular barcodes for soil nematode identification. Mol Ecol 11:839–850

Francis CM, Borisenk AV, Ivanova NV, Eger JL, Lim BK, Guillen-Servent A, Kruskop SV, Mackie I, Hebert PDN (2010) The role of DNA barcode in understanding and conservation of mammal diversity in southeast Asia. PLoS One 5(9):e12575

Galan M, Pagès M, Cosson JF (2012) Next-generation sequencing for rodent barcoding: species identification from fresh, degraded and environmental samples. PLoS One 7(11):e48374

Galimberti A, De Mattia F, Losa A, Bruni I, Federici S, Casiraghi M, Martellos S, Labra M (2013) DNA barcoding as a new tool for food traceability. Food Res Int 50:55–63

Geiser DM, Klich MA, Frisvad JC, Peterson SW, Varga J, Samson RA (2007) The current status of species recognition and identification in Aspergillus. Stud Mycol 59:1–10

Gilmore SR, Gräfenhan T, Louis-Seize G, Seifert KA (2009) Multiple copies of cytochrome oxidase 1 in species of the fungal genus *Fusarium*. Mol Ecol Resour 9(Suppl S1):90–98

Gray JS (1997) Marine biodiversity: patterns, threats and conservation needs. Biodivers Conserv 6:153–175

Greenstone MH, Rowley DL, Heimbach U, Lundgren JG, Pfannenstiel RS, Rehner SA (2005) Barcoding generalist predators by polymerase chain reaction: carabids and spiders. Mol Ecol 14:3247–3266

Gupta SK, Verma SK, Singh L (2005) Molecular insight into a wildlife crime: the case of a peafowl slaughter. Forensic Sci Int 154:214–217
Gupta SK, Bhagavatula J, Thangaraj K, Singh L (2011) Case report. Establishing the identity of the massacred tigress in a case of wildlife crime. Forensic Sci Int Genet 5:74–75
Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006) DNA barcodes distinguish species of tropical Lepidoptera. Proc Natl Acad Sci USA 103:968–971
Hanner R, Becker S, Ivanova NV, Steinke D (2011) FISH-BOL and seafood identification: geographically dispersed case studies reveal systemic market substitution across Canada. Mitochondrial DNA 22(S1):106–122
Hebert PDN, Gregory TR (2005) The promise of DNA barcoding for taxonomy. Syst Biol 54:852–859
Hebert PDN, Cywinska A, Ball SL, de Waard JR (2003) Biological identifications through DNA Barcodes. Proc Biol Sci 270(1512):313–321
Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. PLoS Biol 2:1657–1663
Hebert PDN, de Waard JR, Landry J-F (2010) DNA barcodes for 1/1,000 of the animal kingdom. Biol Lett 6:359–362
Hebert PDN, Ratnasingham S, Zakharov EV, Telfer AC, Levesque-Beaudin V, Milton MA, Pedersen S, Jannetta P, de Waard JR (2016) Counting animal species with DNA barcodes: Canadian insects. Philos Trans R Soc B 371:20150333
Hendrich L, Moriniere J, Haszprunar G, Hebert PDN, Hausmann A, Kohler F, Balke M (2015) A comprehensive DNA barcode database for Central European beetles with a focus on Germany: adding more than 3500 identified species to BOLD. Mol Ecol Resour 15:795–818
Hogg ID, Hebert PDN (2004) Biological identification of springtails (Hexapoda: Collembola) from the Canadian Arctic, using mitochondrial DNA barcodes. Can J Zool 82:749–754
Holmes BH, Steinke D, Ward RD (2009) Identification of shark and ray fins using DNA barcoding. Fish Res 95:280–288
Hou DY, Song JY, Shi LC, Ma XC, Xin TY, Han JP (2013) Stability and accuracy assessment of identification of traditional Chinese materia medica using DNA barcoding: a case study on Flos Lonicerae japonica. Biomed Res Int 2013:549037
Hubert N, Hanner RH, Holm E, Mandrak NE, Taylor EB, Burridge M et al (2008) Identifying Canadian freshwater fishes through DNA barcodes. PLoS One 3:e2490
Hugenholtz P, Goebel BM, Pace NR (1998) Impact of culture independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol 180(18):4765–4774
Hurwood DA, Dammannagoda S, Krosch MN, Jung H, Salin KR, Youssef MA-BH, de Bruyn M, Mather PB (2014) Impacts of climatic factors on evolution of molecular diversity and the natural distribution of wild stocks of the giant freshwater prawn (*Macrobrachium rosenbergii*). Freshw Sci 33(1):217–231
Jo H, Gim J-A, Jeong K-S, Kim H-S, Joo G-J (2014) Application of DNA barcoding for identification of freshwater carnivorous fish diets: is number of prey items dependent on size class for Micropterus salmoides? Ecol Evol 4(2):219–229
Johnsen A, Rindal E, Ericson PGP et al (2010) DNA barcoding of Scandinavian birds reveals divergent lineages in trans-Atlantic species. J Ornithol 151:565–578
Jose D, Nidhin B, Anil Kumar KP, Pradeep PJ, Harikrishnan M (2016) A molecular approach towards the taxonomy of fresh water prawns Macrobrachium striatum and M. equidens (Decapoda, Palaemonidae) using mitochondrial markers. Mitochondrial DNA 27(4):2585–2593
Kermarrec L, Franc A, Rimet F, Chaumeil P, Frigerio J-M, Humbert J-F, Bouchez A (2014) A next-generation sequencing approach to river biomonitoring using benthic diatoms. Freshw Sci 33 (1):349–363
Kerr KCR, Stoeckle MY, Dove CJ, Weigt LA, Francis CM, Hebert PDN (2007) Comprehensive DNA barcode coverage of North American birds. Mol Ecol Notes 7:535–543
Kerr KCR, Lijtmaer DA, Barreira AS et al (2009) Probing evolutionary patterns in Neotropical birds through DNA barcodes. PLoS One 4:e4379

Keskin E, Atar HH (2013) DNA barcoding commercially important fish species of Turkey. Mol Ecol Resour 13(5):788–797

Kim MS, Yang MY, Cho GY (2010) Applying DNA barcoding to Korean Gracilariaceae (Rhodophyta) Cryptogamie. Algologie 31(4):387–401

Kondo T, Gullan PJ, Williams DJ (2008) Coccidolog. The study of scale insects (Hemiptera: Sternorrhyncha: Coccoidea). Revista Corpoica – Ciencia Y Tecnologia Agropecuaria 9:55–61

Kruger A, Obermayr U, Czajka C, Bueno-Mari R, Jost A, Rose A (2014) COI sequencing for invasive mosquito surveillance in Germany reveals genetically divergent specimens near Aedes geniculatus (Diptera: Culicidae). J Eur Mosq Control Assoc 32:22–26

Kumar NP, Rajavel AR, Natarajan R, Jambulingam P (2007) DNA barcodes can distinguish species of Indian mosquitoes (Diptera: Culicidae). J Med Entomol 44:1–7

Kumar NP, Krishnamoorthy N, Sahu SS, Rajavel AR, Sabesan S, Jambulingam P (2013) DNA barcodes indicate members of the Anopheles fluviatilis (Diptera: Culicidae) species complex to be conspecific in India. Mol Ecol Resour 13:354–361

Lakra WS, Singh M, Goswami M, Gopalakrishnan A, Lal KK, Mohindra V, Sarkar UK, Punia PP, Singh KV, Bhatt JP, Ayyappan S (2016) DNA barcoding Indian freshwater fishes. Mitochondrial DNA 27(6):4510–4517

Lambert DM, Baker A, Huynen L, Haddrath O, Hebert PDN, Millar CD (2005) Is a large-scale DNA-based inventory of ancient life possible? J Hered 96:279–284

Laramie MB, Pilliod DS, Goldberg CS (2015) Characterizing the distribution of an endangered salmonid using environmental DNA analysis. Biol Conserv 183:29–37

Lauber CL, Zhou N, Gordon JI, Knight R, Fierer N (2010) Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. FEMS Microbiol Lett 307:80–86

Layton KKS, Martel AL, Hebert PDN (2014) Patterns of DNA barcode variation in Canadian marine Molluscs. PLoS One 9(4):e95003

Les DH, Moody ML, Jacobs SWL, Bayer RJ (2002) Systematics of seagrasses (Zosteraceae) in Australia and New Zealand. Syst Bot 27:468–484

Li J, Han H, Gao Q, Jin Q, Chi M, Wu C, Zhang A (2012) Species identification of Noctuidae (Insecta: Lepidoptera) with DNA barcoding of support vector machine and neighbor-joining method. J Biosafety 21:308–314

Lim NKM, Tay YC, Srivathsan A, Tan JWT, Kwik JTB, Baloğlu B, Meier R, Yeo DCJ (2016) Next-generation freshwater bioassessment: eDNA metabarcoding with a conserved metazoan primer reveals species-rich and reservoir-specific communities. R Soc Open Sci 3:160635

Links MG, Dumonceaux TJ, Hemmingsen SM, Hill JE (2012) The chaperonin-60 universal target is a barcode for bacteria that enables de novo assembly of metagenomic sequence data. PLoS One 7(11):e49755

Little DP (2014) Authentication of Ginkgo biloba herbal dietary supplements using DNA barcoding. Genome 57:513–516

Lohman DJ, Ingram KK, Prawiradilaga DM et al (2010) Cryptic genetic diversity in "widespread" Southeast Asian bird species suggests that Philippine avian endemism is gravely underestimated. Biol Conserv 143:1885–1890

Lorenz JG, Jackson WE, Beck JC, Hanner R (2005) The problems and promise of DNA barcodes for species diagnosis of primate biomaterials. Philos Trans R Soc B 360:1869–1877

Lowenstein JH, Amato G, Kolokotronis S-O (2009) The real maccoyii: identifying tuna sushi with DNA barcodes – contrasting characteristic attributes and genetic distances. PLoS One 4(11): e7866

Lu L, Chesters D, Zhang W, Li G, Ma Y, Ma H, Song X, Wu H, Meng F, Zhu C, Liu Q (2012) Small mammal investigation in spotted fever focus with DNA-barcoding and taxonomic implications on rodents species from Hainan of China. PLoS One 7(8):e43479

Luo AR, Zhang AB, Ho SYW, Xu WJ, Zhang YZ, Shi WF et al (2011) Potential efficacy of mitochondrial genes for animal DNA barcoding: a case study using eutherian mammals. BMC Genomics 12:84

Maas AE, Blanco-Bercial L, Lawson GL (2013) Reexamination of the species assignment of Diacavolinia pteropods using DNA barcoding. PLoS One 8(1):e53889

Makarova O, Contaldo N, Paltrinieri S, Kawube G, Bertaccini A, Nicolaisen M (2012) DNA barcoding for identification of "*Candidatus phytoplasmas*" using a fragment of the elongation factor Tu gene. PLoS One 7(12):e52092

Marie-Stephane T, Mireille O, Serge K (2012) An integrative morphological and moleculardiagnostics for Typhlodromus pyri (Acari: Phytoseiidae). Zool Scr 41:68–78

Marín JC, Saucedo CE, Corti P, González BA (2009) Application of DNA forensic techniques for identifying poached guanacos (*Lama guanicoe*) in Chilean Patagonia. J Forensic Sci 54 (5):1073–1076

Markmann M, Tautz D (2005) Reverse taxonomy: an approach towards determining the diversity of meiobenthic organisms based on ribosomal RNA signature sequences. Philos Trans R Soc B 360:1917–1924

Molnar JL, Gamboa RL, Revenga C (2008) Assessing the global threat of invasive species to marine biodiversity. Front Ecol Environ 6:485–492

Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B (2011) How many species are there on earth and in the ocean? PLoS Biol 9:e1001127

Mouquet N, Devictor V, Meynard CN et al (2012) Ecophylogenetics: advances and perspectives. Biol Rev 87(4):769–785

Naddaf SR, Oshaghi MA, Vatandoost H (2012) Confirmation of two sibling species among Anopheles fluviatilis mosquitoes in South and Southeastern Iran by analysis of cytochrome oxidase I gene. J Arthropod-Borne Dis 6:144–150

Nagoshi RN, Meagher RL, Brambila J (2011) Use of DNA barcodes to identify invasive armyworm Spodoptera species in Florida. J Insect Sci 11:154

Nagy ZT, Sonet G, Glaw F (2012) First large-scale DNA barcoding assessment of reptiles in the biodiversity hotspot of Madagascar, based on newly designed COI primers. PLoS One 7(3): e34506

Naro-Maciel E, Reid B, Fitzsimmons NN, Le M, Desalle R, Amato G (2010) DNA barcodes for globally threatened marine turtles: a registry approach to documenting biodiversity. Mol Ecol Resour 10:252–263

Natalie JWW (2013) Determining the microbial diversity in chaparral soils before and after wildfires through DNA barcoding. Project Number-S1119, California State Science Fair, Project Summary

Neigel J, Domingo A, Stake J (2007) DNA barcoding as a tool for coral reef conservation. Coral Reefs 26:487–499

Newmaster SG, Grguric M, Shanmughanandhan D, Ramalingam S, Ragupathy S (2013) DNA barcoding detects contamination and substitution in North American herbal products. BMC Med 11:222

Ngamskulrungroj P, Gilgado F, Faganello J, Litvintseva AP, Leal AL, Tsui KM, Mitchell TG, Vainstein MH, Meyer W (2009) Genetic diversity of the *Cryptococcus* species complex suggests that *Cryptococcus gattii* deserves to have varieties. PLoS One 4:e5862

Nieukerken EJV, Doorenweerd C, Stokvis FR, Dick SJ, Groenenberg DSJ (2012) DNA barcoding of the leaf-mining moth subgenus Ectoedemias. str. (Lepidoptera: Nepticulidae) with COI and EF1-a: two are better than one in recognizing cryptic species. Contrib Zool 81:1–24

Nwani CD, Becker S, Braid HE, Ude EF, Okogwu OI, Hanner R (2011) DNA barcoding discriminates freshwater fishes from southeastern Nigeria and provides river system-level phylogeographic resolution within some species. Mitochondrial DNA 22(Suppl 1):43–51

Panprommin D, Panprommin N (2017) Assessment of the DNA barcoding for identification of Trigonostigma somphongsi, a critically endangered species in Thailand. Biochem Syst Ecol 70:200–204

Parvathy VA, Swetha VP, Sheeja TE, Leela NK, Chempakam B, Sasikumar B (2014) DNA barcoding to detect chilli adulteration in traded black pepper powder. Food Biotechnol 28:25–40

Parvathy VA, Swetha VP, Sheeja TE, Sasikumar B (2015) Detection of plant-based adulterants in turmeric powder using DNA barcoding. Pharm Biol 53:1774–1779

Patel S, Waugh J, Millar CD, Lambert DM (2010) Conserved primers for DNA barcoding historical and modern samples from New Zealand and Antarctic birds. Mol Ecol Resour 10:431–438

Pauls SU, Blahnik RJ, Zhou X, Wardwell CT, Holzenthal RW (2010) DNA barcode data confirm new species and reveal cryptic diversity in Chilean Smicridea (Smicridea) (trichoptera: hydropsychidae). J N Am Benthol Soc 29:1058–1074

Prendini L (2005) Comment on "Identifying spiders through DNA barcodes". Can J Zool 83:498–504

Rai PS, Bellampalli R, Dobriyal RM, Agarwal A, Satyamoorthy K, Narayana DBA (2012) DNA barcoding of authentic and substitute samples of herb of the family Asparagaceae and Asclepiadaceae based on the ITS2 region. J Ayurveda Integr Med 3:136–140

Rainer J, de Hoog GS (2006) Molecular taxonomy and ecology of *Pseudallescheria*, *Petriella* and *Scedosporium prolificans* (*Microascaceae*) containing opportunistic agents on humans. Mycol Res 110:151–160

Raja HA, Baker TR, Little JG, Oberlies NH (2017) DNA barcoding for identification of consumer-relevant mushrooms: a partial solution for product certification? Food Chem 214:383–392

Rastogi G, Dharne MS, Walujkar S, Kumar A, Patole MS, Shouche YS (2007) Species identification and authentication of tissues of animal origin using mitochondrial and nuclear markers. Meat Sci 76:666–674

Rebijith KB, Asokan R, Kumar NKK, Krishna V, Chaitanya BN, Ramamurthy VV (2013) DNA barcoding and elucidation of cryptic aphid species (Hemiptera: Aphididae) in India. Bull Entomol Res 3:601–610

Remigio EA, Hebert PDN (2003) Testing the utility of partial COI sequences for phylogenetic estimates of Gastropod relationships. Mol Phylogenet Evol 29:641–647

Renaud AK, Savage J, Adamowicz SJ (2012) DNA barcoding of Northern Nearctic Muscidae (Diptera) reveals high correspondence between morphological and molecular species limits. BMC Ecol 12:24

Roe AD, Rice AV, Bromilow SE, Cooke JE, Sperling FA (2010) Multilocus species identification and fungal DNA barcoding: insights from blue stain fungal symbionts of the mountain pine beetle. Mol Ecol Resour 10:946–959

Rossi N, Mantelatto FL (2013) Molecular analysis of the freshwater prawn macrobrachium olfersii (Decapoda, Palaemonidae) supports the existence of a single species throughout its distribution. PLoS One 8(1):e54698

Ruiz-García M, Pinedo-Castro M, Shostell JM (2014) How many genera and species of woolly monkeys (Atelidae, Platyrrhine, Primates) are there? The first molecular analysis of *Lagothrix flavicauda*, an endemic Peruvian primate species. Mol Phylogenet Evol 79:179–198

Sanches A, Perez WAM, Figueiredo MG, Rossini BC, Cervini M, Galetti PM Jr et al (2011) Wildlife forensic DNA and lowland tapir (Tapirus terrestris) poaching. Conserv Genet Resour 3:189–193

Schneider KL, Marrero G, Alvarez AM, Presting GG (2011) Classification of plant associated bacteria using RIF, a computationally derived DNA marker. PLoS One 6(4):e18496

Scorzetti G, Fell JW, Fonseca A, Statzell-Tallman A (2002) Systematics of basidiomycetous yeasts: a comparison of large subunit D1/D2 and internal transcribed spacer rDNA regions. FEMS Yeast Res 2:495–517

Seifert KA, Samson RA, Dewaard JR, Houbraken J, Lévesque CA, Moncalvo JM, Louis-Seize G, Hebert PD (2007) Prospects for fungus identification using *CO1* DNA barcodes, with *Penicillium* as a test case. Proc Natl Acad Sci USA 104:3901–3906

Shander C, Willassen E (2005) What can biological barcoding do for marine biology. Mar Biol Res 1:79–83

Shivji MS, Chapman DD, Pikitch EK, Raymond PW (2005) Genetic profiling reveals illegal international trade in fins of the great white shark Carcharodon carcharias. Conserv Genet 6 (6):1035–1039

Smith MA, Fisher BL, Hebert PDN (2005) DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. Philos Trans R Soc B 360:1825–1834
Smith MA, Poyarkov NA, Hebert PDN (2008a) CO1 DNA barcoding amphibians: take the chance, meet the challenge. Mol Ecol Resour 8:235–246
Smith PJ, McVeagh SM, Steinke D (2008b) DNA barcoding for the identification of smoked fish products. J Fish Biol 72:464–471
Smith MA, Fernandez-Triana JL, Eveleigh E, Gomez J, Guclu C, Hallwachs W, Hebert PDN et al (2013) DNA barcoding and the taxonomy of Microgastrinae wasps (Hymenoptera, Braconidae): impacts after 8 years and nearly 20000 sequences. Mol Ecol Resour 13:168–176
Steinke D, Zemlak TS, Hebert PDN (2009) Barcoding Nemo: DNA-based identifications for the ornamental fish trade. PLoS One 4(7):e6300
Stoeckle MY (2008) Blog: DNA identifies invasive parasitic wasp' in the barcode of life blog. http://phe.rockefeller.edu/barcode/blog/2008/07/07/dna-identifies-invasive-parasitic-wasps/. Accessed 26 Feb 2011
Stoeckle MY, Hebert PDN (2008) Barcode of life: DNA tags help classify animals. Sci Am 298 (10):39–43
Stoeckle MY, Gamble CC, Kirpekar R, Yung G, Ahmed S, Little DP (2011) Commercial teas highlight plant DNA barcode identification successes and obstacles. Nat Sci Rep 1:42
Swartz ER, Mwale M, Hanner R (2008) A role for barcoding in the study of African fish diversity and conservation. South Afr J Sci 104(4):293–298
Sweeney BW, Battle JM, Jackson JK (2011) Can DNA barcodes of stream macroinvertebrates improve descriptions of community structure and water quality? J North Am Benthol Soc 30 (1):195–216
Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G, Brochmann C, Willerslev E (2007) Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Res 35(3):e14
Tack JF, Berghe E, Polk PH (1992) Ecomorphology of Crassostrea cucullata (Born, 1778) (Ostreidae) in a mangrove creek (Gazi, Kenya). Hydrobiologia 247:109–117
Thao L, James T, Alexandra B et al (2013) Bar-coded Enterobacteria: an undergraduate microbial ecology laboratory module. Am J Educ Res 1(1):26–30
Trivedi S, Ghosh SK, Choudhury A (2013) DNA sequence of cytochrome c oxidase subunit 1 (COI) region of an oyster Saccostrea cucullata collected from Sunderbans. J Environ Sociobiol 10(1):77–81
Trivedi S, Affan R, Alessa AHA et al (2014) DNA barcoding of Red Sea fishes from Saudi Arabia – the first approach. DNA Barcodes 2:17–20
Uchimura M, Faye EJ, Shimada S, Inoue T, Nakamura Y (2008) A reassessment of Halophila species (Hydrocharitaceae) diversity with special reference to Japanese representatives. Bot Mar 51:258–268
Udayasuriyan R, Saravana Bhavan P, Vadivalagan C, Rajkumar G (2015) Efficiency of different COI markers in DNA barcoding of freshwater prawn species. J Entomol Zool Stud 3(3):98–110
Vargas SM, Araujo FCF, Santos FR (2009) DNA barcoding of Brazilian sea turtles (Testudines). Genet Mol Biol 32(3):608–612
Vassou SL, Kusuma G, Parani M (2015) DNA barcoding for species identification from dried and powdered plant parts: A case study with authentication of the raw drug market samples of Sida cordifolia. Gene 559:86–93
Vialle A, Feau N, Allaire M, Didukh M, Martin F, Moncalvo JM, Hamelin RC (2009) Evaluation of mitochondrial genes as DNA barcode for *Basidiomycota*. Mol Ecol Resour 9(Suppl S1):99–113
Wang J, Soininen J, He J, Shen J (2012a) Phylogenetic clustering increases with elevation for microbes. Environ Microbiol Rep 4(2):217–226
Wang G, Li C, Guo X, Xing D, Dong Y (2012b) Identifying the main mosquito species in china based on DNA barcoding. PLoS One 7:e47051
Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. Philos Trans R Soc Lond B 360:1847–1857

Ward RD, Holmes BH, White WT, Last PR (2008) DNA barcoding Australasian chondrichtyans: results and potential uses in conservation. Mar Freshw Res 59(1):57–71
Waycott M, Duarte CM, Carruthers TJB et al (2009) Accelerating loss of seagrasses across the globe threatens coastal ecosystems. Proc Natl Acad Sci USA 106:12377–12381
Weigt LA, Baldwin CC, Driskell A (2012) Using DNA barcoding to assess Caribbean Reef fish biodiversity: expanding taxonomic and geographic coverage. PLoS One 7(7):e41059
Wilson EO (ed) (1994) Biodiversity. National Academy Press, Washington, DC
Wilson EO (2003) The encyclopaedia of life. Trends Ecol Evol 18:77–80
Wong EHK, Hanner RH (2008) DNA barcoding detects market substitution in North American seafood. Food Res Int 41:828–837
Wong KL, But PPH, Shaw PC (2013) Evaluation of seven DNA barcodes for differentiating closely related medicinal Gentiana species and their adulterants. Chin Med 8:16
Woodcock TS, Boyle EE, Roughley RE, Kevan PG, Labbee RN, Smith ABT, Goulet H et al (2013) The diversity and biogeography of the Coleoptera of Churchill: insights from DNA barcoding. BMC Ecol 13:40
Wu H, Wan Q-H, Fang S-G, Zhang S-Y (2005) Application of mitochon-drial DNA sequence analysis in the forensic identification of Chinese sika deersubspecies. Forensic Sci Int 148:101–105
Yang Q, Zhao S, Kucerova Z, Stejskal V, Opit G, Qin M, Cao Y, Li F et al (2013) Validation of the 16S rDNA and COI DNA barcoding technique for rapid molecular identification of stored product Psocids (Insecta: Psocodea: Liposcelididae). J Econ Entomol 106:419–425
Zhou J, Wang W, Liu M, Liu Z (2014) Molecular authentication of the traditional medicinal plant Peucedanum praeruptorum and its substitutes and adulterants by DNA – barcoding technique. Pharmacogn Mag 10:385–390
Zimmermann J, Glöckner G, Jahn R, Enke N, Gemeinholzer B (2015) Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. Mol Ecol Resour 15(3):526–542

"Significance of DNA Barcoding in Avian Species: Tracing the History and Building the Future"

Farhina Pasha

Abstract Avian species or subspecies have almost identical barcode of life; therefore when a new unidentified sample is encountered, all it takes for its recognition is a plumage. Although a feather is dead keratin, but at the end of this feather, some dead skin cells are attached which are the source for DNA extraction and further COI gene sequence amplification. The faecal sample from a fleeting bird having lots of intestinal epithelial cells forms a great sample for DNA extraction. The DNA barcode so obtained can be then compared to the databases available publicly, such as ABBI/BOLD, etc. Despite all the milestones DNA barcoding has achieved, there are certain issues or limitations which must be clearly recognized and resolved. Furthermore, there is a need to conduct extensive studies in tropical taxa and those with limited dispersal. These studies must be undertaken with broader taxonomic lineage and wider geographic boundaries to discover all sister taxa as well.

Keywords DNA barcoding · BOLD · COI · Avian species · Moa · Bird strike

1 Introduction

Fossils are not only interesting discoveries, but they also allow to peep into the past biodiversity, and evolutionary changes occur in that species. There are many techniques to identify the species and the changes which occurred in the species. "DNA barcoding", the much talked technique which has marked its existence in the world of taxonomy and identification of species, using single mtDNA gene cytochrome *c* oxidase I (COI), has partially been appreciated and partially criticized for the same reason that identification is based on a single gene. Main characteristics of DNA barcoding include as follows: (a) The technique uses fragments of DNA for identification or characterization of any species. (b) It can be used for all stages of life and life forms, i.e. the technique can be performed on the cells directly collected from the

F. Pasha (✉)
Department of Biology, Faculty of Sciences, University of Tabuk, Tabuk, Saudi Arabia
e-mail: fbasha@ut.edu.sa

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_4

organism or can be done on the fossils. (c) It unveils the look-alike species. (d) The process is fast and helps in the recognition of known species and detection of new ones and can recognize the relation between two species. (e) It forms standardized DNA barcode libraries which makes it feasible to identify species: native or invasive, rare or abundant or endangered or nascent. It can be well phrased that "DNA barcoding has proliferated new leaves of hopes to the tree of life".

2 Significance of DNA Barcoding in *Aves*

DNA barcoding is not only useful to identify or categorize the organisms, but it can also be conducted on the fossils having traces of DNA. DNA of the fossil or extinct organisms can be collected from the museums all around the world. The museums are full of specimens which have preserved birds' skin and feather and that of other animals too. These can be of significant importance as DNA extraction samples. By using DNA barcoding, the correlation can be established between the birds from the past and the birds from the modern era. This hopefully will solve the cryptic species identification and of course trace original species taxonomic position.

Earlier the sequence diversity in a 648-bp region of the mtDNA, cytochrome *c* oxidase I (COI), was used for the DNA barcode to identify the animal species. This study tested the effectiveness of a COI barcode in establishing the bird species as well. They tested 260 avian species from North America and found that all the 260 avian species sampled had a unique COI sequence and 130 species had two or more than two specimens present as observed during testing, which had its confirmation as their COI were similar or identical to other species. The COI variation amongst the species was 0.43% whereas between the species was reported to be 7.93%. Therefore their results established a positive wave for DNA barcoding technique as a milestone in identification of new species as well as establishing more precise and accurate taxonomy. They also anticipated a "standard sequence threshold" called as "barcoding gap" which was ten times the mean intraspecific variation for the group under study (Hebert et al. 2004).

3 Indexing the Ancient Biota

DNA barcoding can be used for the identification of different species of plants as well as animals of the present world. It also helps to give insight of the past species and their correlation with the present. DNA barcoding can be conducted on the DNA isolated from the fossils. In an effort to decipher the ancient life via DNA barcoding, Lambert et al. (2005) sequenced 26 subfossil moa bones by COI gene analysis. The result showed a unique barcode and intraspecific COI sequence with a variance range of 0–1.24%. To decipher the novel species, a standard sequence threshold of 2.7% COI sequence was set. The above results were in confirmation with that of

results. Employing this value six different moa groups were identified. With the standard sequence threshold of 1.24%, ten moa species were identified amongst the known group with one deviation. This probably was a previously unidentified species. It is also inferred that due to slow rate of growth and reproduction, the interspecies variation was also very slow.

In other study, Ewan Grant-Mackie (2006), a mathematician, determined the DNA barcodes of pūkeko, takahē, moho, Samoan swamp hens, Solomon Island swamp hens, Tongan swamp hens and Australian swamp hens from North Island and South Island. The samples of moho and Polynesian (Samoan, Solomon Island and Tongan) swamp were collected in 1920. He conducted the DNA barcoding of the mitochondrial cytochrome *c* oxidase subunit 1 (CO1) gene. The DNA barcoding results indicated that the North Island pukekos were of Australian origin and those birds came to North Island by flying and the South Island pukekos were from Polynesian origin.

Avian species or subspecies have almost identical barcode of life; therefore when a new unidentified sample is encountered, all it takes for its recognition is a plumage. The feather is dead keratin, but in the end of this feather, some dead skin cells are attached which are the source of DNA traces and COI gene sequence amplification. Not only feathers but also the faecal sample from a fleeting bird having lots of intestinal epithelial cells forms a great sample for DNA extraction. The DNA barcode so obtained can be then compared to the databases available publicly: such as ABBI/BOLD, etc. Similarly if the fossils of some ancient birds are available, then the DNA can be extracted from the bone samples which can be used for DNA barcoding. Methodology for analysis of avian samples is exhibited in Fig. 1.

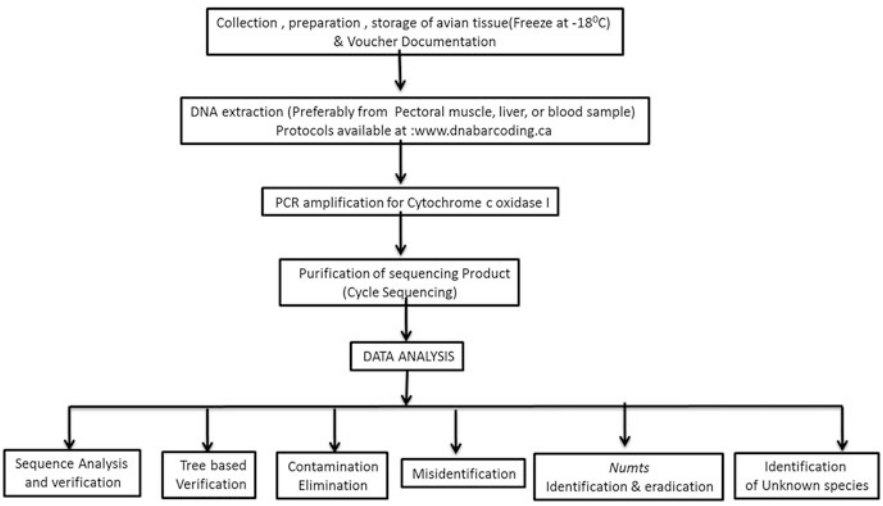

Collection , preparation , storage of avian tissue(Freeze at -18°C)
& Voucher Documentation
DNA extraction (Preferably from Pectoral muscle, liver, or blood sample)
Protocols available at :www.dnabarcoding.ca
PCR amplification for Cytochrome c oxidase I
Purification of sequencing Product
(Cycle Sequencing)
DATA ANALYSIS
Sequence Analysis and verification
Tree based Verification
Contamination Elimination
Misidentification
Numts Identification & eradication
Identification of Unknown species

Fig. 1 Modus operandi—sampling in avian species for DNA barcoding

4 The Present Scenario

The success of this technique exhibited that there are over 198 publications from 2004 to 2017. In a study, a desert wheatear spotted on the Scottish Coast in winter 2013 had its origin in Gobi Desert; a yellow wagtail found around the sewage farm in northeast Russia and a black redstart which was identified by its faecal matter left on the beachfront during migration all came from China (Martin 2013).

In another study ornithologist Craig Symes closely observed a white-winged flufftail (*Sarothrura ayresi*). White-winged flufftail is so rare that it has been spotted in 15 sites only in South Africa in the past 136 years. Ornithologists believe that there are only 250 left, so the bird comes under critically endangered species. Very little information is available about this bird, and it is still not clear whether they are two different species in South Africa and Ethiopia or they have migrated from one place to another as this small water bird with big feet was also spotted in Ethiopia, the only country after South Africa (Dalton et al. 2016). After careful genetic and isotope analysis, Craig and his team (2013) confirmed the two birds are of the same species. Further DNA barcode studies are underway.

Dove et al. (2008) used the technique of using mitochondrial DNA barcodes (*cytochrome c oxidase subunit 1* [CO1]) for identifying the birds which are the major cause of air collision. As the sample available after collision is very small, molecular analysis is not possible; therefore DNA barcoding has emerged as an immensely useful technique. A total of 821 samples were collected from September to December 2006 from the United States for DNA analysis. From the total sample collected, 554 (67.5%) of the sample was successfully amplified for DNA barcoding, whereas 267 (32.5%) could not be amplified. The age of the samples which was almost 6 months did not affect the DNA efficiency of the sample. It was observed that primary state of the sample and its collection method greatly influences the DNA viability. Related studies were performed by many groups to prove this fact (Blackwell and Wright 2006; Christidis et al. 2006; Dolbeer 2006; Doran et al. 1990; Hebert et al. 2003; Hermans et al. 1996; Laybourne and Dove 1994; Linnell et al. 1996; Yoo et al. 2006).

5 Limitations of DNA Barcoding

Despite all the milestones DNA barcoding has achieved, there are certain issues or limitations which must be clearly recognized and resolved. Some of these mentioned below not only related to biological aspects but also effect statistical status.

5.1 Elimination of "Barcode Gaps"

Barcode Gaps may be due to shortage of sampling or sampling quality (Meyer and Paulay 2005). Careful assessment of sampling during database assembly phase is a must to avoide the Barcode Gaps (Wiemer and Fiedler 2007). Meyer and Paulay (2005) also observed the inaccuracies for some specimens in well-defined phylogenies as well as in moderately identified groups.

5.2 Intrinsic Risks Due to Mitochondrial Lineage

As a well-known fact that the mitochondrial DNA (mtDNA) is strongly related to maternal inheritance, the usage of mt loci many times leads to pseudo-estimation of sample divergence thereby leading to an ambiguous identification. Some processes like interspecific hybridization and endosymbiont infections may lead to relocation of mitochondrial gene group (Dasmahapatra and Mallet 2006). In all the above conditions, nuclear loci analysis is required to clear the phylogenetic relationships. These should be clearly marked in BOLD during its database entry.

5.3 NUMTs Nuclear Copies of COI

This is another issue to be cautiously addressed in DNA barcoding. Nuclear mitochondrial DNAs (NUMTs) are described as nuclear copies of mitochondrial DNA sequences, which are translocated into the nuclear genome (Williams and Knowlton 2001). The range of NUMTs varies with species, e.g. none or few in *Anopheles* to more than 500 in humans (Richly and Leister 2004). NUMTs can be predicted during sequence or amino acid alignment. They can be traced by the sequence checking method available in BOLD, namely, "rejection of inconsistent amino acid alignment".

5.4 Rate of Genomic Evolution

The rate with which species evolve is not same for all species. There have been cases where due to the lack of COI sequence, resolving power has mislead taxa from primary single-gene method to multiple region barcoding system (especially when COI sequence is non-species specific or in taxa with low mitochondrial evolutionary rate). This is termed as NON-COI barcode (Bakker 2007).

5.5 *Geographical Structure*

Particularly in the case of avian species, geographical structure can smudge species delineation, as elevated rates of intraspecific divergence are obtained from geographically remote or isolated population (Hebert et al. 2003) and therefore must be a major consideration in DNA barcoding. To resolve a prominent question of boundary between population and species, we need to incorporate broad-range intraspecific samples in the reference databases. Furthermore, there is a need to conduct extensive studies in tropical taxa and those with limited dispersal. These studies must be undertaken with broader taxonomic lineage and wider geographic boundaries to discover all sister taxa as well.

6 Future Prospects

Birds are the best example for the study on the taxonomy and phylogenetic level than any other organism. Birds are probably one of the best groups of vertebrates in frozen tissue collections, with more than 300,000 tissue samples covering nearly 75% of known bird species (Stoeckle and Winker 2009). This specificity of birds makes them the ideal for the analysis of the effectiveness of any optimized genetic method for species identification or in simple words DNA barcoding. Aves are the first group on which the large-scale DNA barcoding studies were conducted (Lijtmaer et al. 2012). Birds are found to be present in almost all the habitat and show sensitivity to the change in the environmental conditions. Therefore they play important role in the environmental monitoring as indicator species (Wormworth). DNA barcoding is one of the simple and reliable methods to identify the different species of the aves of past, present and future and their correlation with the environmental changes. Despite the simplicity and reliability of DNA barcoding, there are a few disadvantages as the technique relies only on mtDNA single-gene analysis. DNA barcoding can be more reliable and useful if it would be supplemented with other barcodes especially from the nuclear barcodes (Dasmahapatra) and the carbon dating for the fossils. There are many recent advances in barcoding techniques; one of them is metabarcoding. Metabarcoding is a fast and quick method to assess the biodiversity. The method is the combination of DNA barcoding (where the universal PCR primers are used to mass-amplify DNA barcodes) and high-throughput DNA sequencing.

As the proverb stands “and miles to go before I sleep”, DNA barcode has still to solve many more mysteries by systematizing millions of sequences being generated every day in laboratories all over the world.

7 Conclusion

To determine the time of diversification of modern birds is a difficult task. Many methods that are required to determine the different DNA sequencing studies reveal that the ancestors of modern birds were inhabiting in South America 95 million years ago (Claramunt and Cracraft 2015). DNA barcoding is one of the powerful tools to determine the age and the origin of a bird species. It is efficient in displaying the avian realms and demarcating their taxonomic boundaries. DNA barcoding is also helpful in clearing the ambiguity for point of origin of an avian species from their point of existing habitat. Although there are few drawbacks of using DNA barcoding, but the success of barcoding to distinguish species in a wide range of taxa and to identify the cryptic species is remarkable (Schirtzinger et al. 2012). It can be performed on the small quantity of samples collected from the fossils, which means the fossil's DNA barcoding can give insight of the origin of bird species when it combines with other techniques.

Acknowledgements The authors would like to acknowledge University of Tabuk, Tabuk, Saudi Arabia.

The author would also like to thank the Department of Biology, Faculty of Sciences, Saudi Digital Library and University Library for providing the facility for literature survey and collection.

References

Bakker (2007) "Second International Barcode of Life" Conference, TAIPEI

Blackwell BF, Wright SE (2006) Collisions of red-tailed hawks (*Buteo jamaicensis*), turkey vultures (*Cathartes aura*), and black vultures (*Coragyps atratus*) with aircraft: implications for bird strike reduction. J Raptor Res 40:76–80

Christidis L, Norman JA, Johnson RN, Lindsay S (2006) Forensic identification of aviation bird strikes in Australia. Australian Government, Australian Transport Safety Bureau Research and Analysis Report, Aviation Safety Research Grant B2005/0017. Australian Museum, Canberra

Claramunt S, Cracraft J (2015) A new time tree reveals Earth history's imprint on the evolution of modern birds. Sci Adv 1(11):e1501005

Dalton DL, Vermaak E, Smit-Robinson HA, Kotze A (2016) Lack of diversity at innate immunity toll-like receptor genes in the critically endangered white-winged Flufftail (*Sarothrura ayresi*). Sci Rep 6

Dasmahapatra KK, Mallet J (2006) DNA barcodes: recent successes and future prospects. Heredity 97:254–255

Dolbeer RA (2006) Height distribution of birds recorded by collisions with civil aircraft. J Wildl Manag 70:1345–1350

Doran HJ, Cross TF, Kelly TC (1990) Electrophoretic identification of bird species involved in collisions with aircraft. Comp Biochem Physiol 97:171–175

Dove CJ, Rotzel NC, Heacker M, Weigt LA (2008) Using DNA barcodes to identify bird species involved in bird strikes. J Wildl Manag 72(5):1231–1236. https://doi.org/10.2193/2007-272

Ewan Grant-Mackie (2006) http://legacy.biotechlearn.org.nz/themes/barcoding_life/barcoding_new_zealand_swamp_hen

Hebert P, Cywinska A, Ball SL, de Waard JR (2003) Biological identification through DNA barcodes. Proc R Soc Lond B 270:313–321

Hebert PD, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. PLoS Biol 2(10):e312

Hermans J, Buurma L, Wattel J (1996) Identification of bird remains after bird–airplane collisions, based on DNA sequences analysis. Proc Int Bird Strike Committee Eur 23:203–207

Lambert DM, Baker A, Huynen L, Haddrath O, Hebert PD, Millar CD (2005) Is a large-scale DNA-based inventory of ancient life possible? (PDF fulltext). J Hered 96(3):279–284

Laybourne RC, Dove CJ (1994) Preparation of bird-strike remains for identification. Proc Bird Strike Committee Eur 22:5531–5543

Lijtmaer DA, Kerr KC, Stoeckle MY, Tubaro PL (2012) DNA barcoding birds: from field collection to data analysis. In: DNA barcodes: methods and protocols, pp 127–152

Linnell MA, Conover MR, Ohashi TJ (1996) Analysis of bird strikes at a tropical airport. J Wildl Manag 60:935–945

Martin C (2013) CSI: Birding … barcoding the past, present and future; the conservation. Academic rigour, journalistic flair

Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. PLoS Biol 3:2229–2238

Richly E, Leister D (2004) NUMTs in sequenced eukaryotic genomes. Mol Biol Evol 21:10811084

Schirtzinger EE, Tavares ES, Gonzales LA, Eberhard JR, Miyaki CY, Sanchez JJ, Hernandez A et al (2012) Multiple independent origins of mitochondrial control region duplications in the order Psittaciformes. Mol Phylogenet Evol 64(2):342–356

Stoeckle M, Winker K (2009) A global snapshot of avian tissue collections: state of the enterprise. Auk 126:684–687

Wiemer M, Fiedler K (2007) Does the DNA barcoding gap exist? – a case study in blue butterflies (Lepidoptera: Lycaenidae). Front Zool 4:8. https://doi.org/10.1186/1742-9994-4-8

Williams ST, Knowlton N (2001) Mitochondrial pseudogenes are pervasive and often insidious in the snapping shrimp genus Alpheus. Mol Biol Evol 18:1484–1493

Yoo HS, Eah J, Kim JS, Min YKM, Paek WK, Lee H, Kim C (2006) DNA barcoding Korean birds. Mol Cells 22:323–327

DNA Barcoding: A Potential Tool for Invasive Species Identification

Muniyandi Nagarajan, Akash Nambidi Parambath, and Vandana R. Prabhu

Abstract Invasive alien species epitomize the group of nonindigenous species that invade a geographic area and impose detrimental effects environmentally, ecologically and economically. The management of invasive species will be less complicated at the primary stages of invasion, since their population will be represented by a few numbers of individuals. But the detection of these few individuals will be a strenuous task if they are cryptic and exhibits low detectability. Molecular diagnostic tools bestow favourable support for the precise and brisk detection of morphologically indefinite alien species. The more recent outbreak of DNA barcoding lends as an aspiration for the assessment of biodiversity in a more accurate and also in an inexpensive manner. This typical procedure stands out as a reliable toolkit for the detection of individual/bulk samples, forensic residues as well as environmental DNA and enhances rapid response management. Several DNA barcodes including mitochondrial COI gene, rbcL, matK, trnH-psbA, ITS (nuclear internal transcribed spacer), etc., have been extensively used as a global bio-identification system for detecting the alien species that invade different ecosystems. This chapter confers about invasive species and the implication of DNA barcoding in their identification as well as management process.

Keywords Invasive species · DNA barcoding · COI gene · Identification · Management

1 Introduction

Biodiversity, which is an elision of biological diversity, acts as the bedrock of the structure and function of an ecosystem and thereby contributes to the balance of nature. The ecosystem balance is disturbed by the small-scale risks associated with

M. Nagarajan (✉) · A. N. Parambath · V. R. Prabhu
Department of Genomic Science, School of Biological Sciences, Central University of Kerala, Kasargod, Kerala, India
e-mail: nagarajan@cukerala.ac.in

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_5

the diverseness that detonates rapidly, and thus, emergence of potent policies like biosecurity comes into sight. These small-scale risks include the problems associated with infectious disease, genetically altered organisms, biological weapons like a number of biological agents that are used as a weapon against humans, plants or animals and the ways by which all these can be minimized (Meyerson et al. 2002). It also involves consequential role of biological invasions that throws threat to the medley of life forms. According to the Convention on Biological Diversity (CBD), invasive alien species (IAS) is the species whose establishment outside their natural habitat threatens ecosystem, habitat or species causing infliction to both environment and economy. IAS pressurizes the ecosystem stability, producer sustenance and consumer confidence (Cock et al. 2003). They cause unalterable detrimental effects on biodiversity, like loss of habitat, followed by the extinction of native species by replacing them with invasive species. Since agricultural and commercial activities have an inevitable dependence on our natural ecosystem, the undesirable impact of invasive species costs billions of dollars to the economy. Breaching of biogeographic barriers and the influx of exotic species all over the world through increased international trade and tourism, climate change, etc., shows a substantial part in serious economic impacts. For example, IAS stands as the paramount of environmental damages and the cost of these losses add up to almost $120 billion per year in the Unites States. Also, about 42% of the endangered species are at risk predominantly because of invasive species (Pimentel 2005).

Over the centuries, introduction of exotic species has been done intentionally or unintentionally by anthropogenic actions. These exotic species established rapidly in those regions, causing major economic and environmental issues. Thus, in an ecosystem, it is more desirable to prevent the ignition of invasive species before they become more established, because that would be more cost-effective and secure. To contend this, several management measures have to be taken like forbidding the introduction and establishment of alien species, early detection and annihilation of exotic species (Simberloff et al. 2005). The most significant part in monitoring and predicting a new specimen is to accurately distinguish the specimen to the species level. It requires collection of large amount of accurate data about the species entering non-native habitat. However this comprehensive identification process is interrupted by the non-availability of a taxonomic expert. Also, the existence of morphologically undistinguishable invasive species makes the detection procedure tedious and leads to the demand for a rapid and accurate identification technique for their detection. This leads to the unravelling of DNA-based tools for identifying and monitoring the invasive species.

DNA-based molecular tools guarantee to enhance over traditional monitoring approaches by extending the detection sensitivity and faster performance. Among the different DNA-based molecular tools, DNA barcoding has been proposed as the standardized universal method for monitoring and identification of different species (Hebert et al. 2003). Barcoding uses a very short DNA sequence from a standard part of the genome which is common across all taxa. Thus it stands out of several other molecular diagnostic tools because it can serve a dual purpose, both acting as a taxonomist's toolbox enriching their knowledge as well as being an innovative

device for non-experts who need to make a quick identification. Variation between two organisms is known as the barcoding gap (Meyer and Paulay 2005). DNA barcoding enables rapid identification of the specimens at any life stage (Palumbi and Cipriano 1998). This success of DNA barcoding in species identification is dependent mainly on the accuracy of a reference database, which consists of voucher specimens that are morphologically identified using existing methods and compared to the sequence of a given barcode (Ratnasingham and Hebert 2007). The main motive of DNA barcoding is to parent the global BOLD database, which includes the sequence of voucher specimen as well as the sequence imported from other database.

2 Invasive Species: A General Outlook

The species architecture of an ecosystem depends mainly on environmental conditions, nature of disturbances and counterbalance of extinction and recruitment. The disturbances caused by invasive species are one among the major types of disturbances that alters the species composition. Invasive species are benefited if they are unaccompanied by their natural enemies. This ecological release helps them to get established in the new habitat to which they are introduced. Each species has distinctive potential for becoming invasive, and to study about this, knowledge about their rare dispersal events along long distances are required. Other factors such as habitat disturbances, distribution frequency and reproductive maturity also play a crucial role in determining the invasive potential of different species. Among 2 million species that are characterized, almost 10% (200,000) species have the ability to become cogent invaders. Some examples of IAS that belong to different categories are given in Table 1.

3 Impacts of Invasive Alien Species

The establishment of invasive species can cause harm to the native species, economy, environment and human health. Consequences of invasion can be dramatic. It can also alter the evolutionary pathway of native species and ultimately destroy the ecological balance by species extinction. Even though the species transported across biogeographic barriers increases, species richness does not show much variation. Only few of them got established and about 1% of them become pests. But over the years, these additions have become significant (Williamson 1996). In New Zealand alien established plant species equals to the proportion of native plant species. Most of the countries contain 20% or more alien plant species (Vitousek et al. 1996). Vast lands and water capes in different regions are dominated by invasive alien species like star thistle (*Centaurea solstitialis*) in California and cheat grass (*Bromus tectorum*) in the United States (Mack 1985), all of which disturbs the balance of local biodiversity.

Table 1 Different types of invasive alien species

Organisms	Descriptions
Avian malaria (*Plasmodium reticulum*)	It was introduced to Hawai'i by means of some exotic birds, which was further spread in 1826 by a vector named southern house mosquito (*Culex quinquefasciatus*) that occupied the water barrels in a sailing ship. Unlike non-native birds, native birds of Hawai'i do not have resistance against avian malaria, and their number declined quickly. Birds such as honeycreepers which have evolved into a diverse array of species are mostly affected by this virus. Avian malaria has led to the extinction of at least ten native bird species and terrorizes many more in the islands of Hawai'i
Water hyacinth (*Eichhornia crassipes*)	It is native to South America and considered as the most dreadful aquatic weed in the world. It has gotten large purple and violet-coloured flowers which make them visually attractive and popular ornamental plant. It is now commonly seen in almost 50 countries as it is a very fast-growing aquatic weed that doubles its population size within 12 days. It rapidly spreads all over the waterways making it difficult for swimming and fishing and boat traffic. Moreover, it prevents sunlight and oxygen from reaching water column that results in the dramatic reduction of aquatic life forms and eventually changes the entire aquatic ecosystem
Miconia (*Miconia calvescens*)	A native of South America and it is a highly ornamental plant. Their red and purple leaf makes them very attractive. It was imported to the botanical garden on the island of Tahiti in 1937. Currently more than half of the island is invaded by this plant, being spread by fruit-eating birds making it difficult for the survival of endemic species. It was also introduced to islands in Hawai'i as an ornamental plant, where its population is still increasing
Crazy ant (*Anoplolepis gracilipes*)	They are invaders seen on Christmas Island in Indian Ocean. They have caused many environmental problems to natural ecosystems which include the extermination of land crab (*Gecarcoidea natalis*) population. These invaders attack a variety of arthropods, birds and mammals and are also involved in the protection of sap-sucking scale insects that destroy forest canopy. They have invaded almost 5% of Christmas Island, and if this continues it will lead to the extinction of many species on the island
Bullfrog (*Lithobates catesbeianus*)	It was introduced to northern Belgium in the 1990s along with fish transports from other European countries. Later it spreads to other parts of the country because of the availability of suitable reproductive conditions like the nutrient-rich ponds, lack of predators and presence of lots of algae
Western mosquitofish (*Gambusia affinis*)	It is a small fish native to the Southern United States. It was introduced to different parts of the world as a biological control of mosquito mainly by mosquito control agencies without knowing its adverse effects. It has become a pest worldwide because it eats the eggs of rare and economically desirable fish
The house sparrow (*Passer domesticus*)	It is native to Asia and introduced to Eastern Africa on trading ships long ago. Even though they mainly feed on insects, it can cause harm to humans by destroying fruit trees, crops and roof tops by its nesting activity. Introduction of this species also led to the

(continued)

Table 1 (continued)

Organisms	Descriptions
	decrease in the number of several native species by taking over its nesting place
Brown tree snake (*Boiga irregularis*)	It is native to Australia, Indonesia, Papua New Guinea and the Solomon Islands. It has been introduced to Guam on military aircrafts in the 1940s. Their population rate explode as they were unaccompanied by their natural predators which led to serious economic and ecological destruction. It is involved in the complete extermination of Guam's native bird species. It has also entered Hawai'i, the United States, Spain and many others by concealing itself in boats, aircrafts and airplane wheel wells causing serious threat to the biological diversity
Feral pig (*Sus scrofa*)	It was once domesticated but has been released or escaped into the wild. Subsequently it spread to different parts of the world like Australia, Canada, the United States and many others. They cause serious threat to native vegetation including humans by damaging crops and transmitting disease like leptospirosis

As global trade increases, the rate of exchange of exotic species also increases. Some examples from the past also show dominant nature of invasive species in their non-native habitat. The biota of the Red Sea and Mediterranean Sea were isolated until the construction of Suez Canal in 1869. Suez Canal opened the way for movement of different species across the Red Sea and Mediterranean Sea. Over 250 species moved into the Mediterranean Sea from the Red Sea (Por 1978). This invasion resulted in the displacement of native fish to the depths as the invaders preferred shallow, warmer water at surface (Golani 1993). This helped the invaders to exploit the new habitat and increase in number. As their population make greater in size, they had greater negative impact on ecosystem. The impacts of anthropogenic actions like fire or pollution decrease over time, but impacts of IAS tend to increase over time. The impacts of IAS can be mainly categorized as follows.

3.1 Ecological Impacts

Alteration of the local biodiversity of an area or ecological process by invasive species can cause many ecological impacts. Their impacts can be seen at different level starting from individual species to local population that disturb the function of the entire ecosystem. For example, invasive species can hybridize with the native species causing alteration in the gene pool resulting in reduced fitness of native species, or alternatively invaders become more stable. Hybridization event eventually leads to the establishment and propagation of invasive species. Invasive species particularly plants are capable of changing the structure of soil, causing erosion problems and altering resource availability like minerals and water. This in turn leads to cascade of negative impacts on ecosystem. For example, the paper bark tree

(*Melaleuca quinquenervia*), an invasive species introduced to Florida, resulted in the alteration of soil structure and resource availability. Subsequently the wetlands in this area are subjected to degradation (Porazinska et al. 2007). Likewise invasion of cheat grass (*Bromus tectorum*) in the Northwestern United States enhanced the fire frequency by increasing the habitat of cheat grass (Mack 1985).

3.2 Economic Impacts

Invasive species may cause crucial economic losses to the society either in the form of direct economic impacts like loss of crops due to invasive crop pests, preventing export of products that are infected by IAS, etc., or indirect or secondary impacts such as human health issues, cost of responding and prevention of the problems caused by IAS, damage to infrastructure due to ecosystem changes, etc., which stand foremost in the list of economic impact contributed by IAS.

New Zealand ranks as one of the most highly invaded areas in the world due to its increased number of introduced mammals, birds and several exotic plant species. Some among those species that have been introduced for prosperity in the field of agriculture, forestry, horticulture, etc. was found to engender threats to native biodiversity. For the last 150 years, these invading species are found to be naturalizing at a fixed rate of 12 per year. The New Zealand economy have a loss of about 400 million NZ dollars in a year due to exotic species, and also about 440 million NZ dollars cost for the prevention of these losses. The sum of these two costs contributes to 1% of New Zealand's GDP (Williams and Timmins 2002). The estimated annual cost for damages and losses rendered by the domestic cats in the United States was about 17 billion dollars, which only represents the damage done by feral cats to certain bird species. The total loss triggered by both feral and urban (pet) cats together amasses around 34 billion dollars (Pimentel et al. 2005). It is inauspicious that a single invasive alien species, the water hyacinth (*Eichhornia crassipes*), cost about several millions of US dollar loss in the United States from 1980 to 1991. For instance, in Florida more than 43 million US dollars are spent to counteract and another 3 million dollars for the management of this cryptic species every year. 500,000 US dollars are spent annually in California for the control of this weed (Mullin et al. 2000).

3.3 Impact on Human Health

Apart from creating loss to the biodiversity, invasive species also cause menace to human health. Invasive pathogens can affect health directly, or, otherwise, invasive vectors can sometimes alter the transmission cycles of native or non-native pathogens (McMichael and Bouma 2000). Many of the non-native species, including

insects, rodents and birds, can act as reservoirs for disease and can even carry diseases such as yellow fever and malaria.

3.4 Consequences of Species Mixing

It has been studied that invasive species can adapt to their new environmental conditions pretty quickly which is known as the direct evolutionary consequence of species mixing. Huey et al. (2000) demonstrated evolution of fruit fly, which took only 20 years to change its wing size to get adapted to the new habitat, west coast of North America. Even though the developmental basis for the change in wing size was different from that of native European populations, functional result was the same. On the other hand, there are studies that show the evolution of native species towards introduced species (Carroll and Dingle 1996). In addition to direct responses, there are many indirect responses towards species mixing. The major causes for these responses are hybridization and introgression. Rhymer and Simberloff (1996) summarized that hybridization of invasive species with native species can cause loss of fitness in the native species. This may even lead to the extinction of native species.

4 Traditional Methods and Identification of Invasive Species

Exploration and classification of living organisms became an essential process throughout history. The need for recognition of unknown species varies, and they are spread across different fields like forensic studies, conservational biology, agriculture and so on. Traditional methods provide fundamental information for the assessment of unknown species. It basically works by collecting necessary information about the uniqueness, locality and prosperity of alien species. These information are necessary for the management of invasive species, which are one among the growing threats of the ecosystem. Traditional methods mainly rely on morphological characters for the identification process. They are mainly of two types: one that documents all the vital information about invasive species including pattern of distribution and abundance, which are necessary for their management process, and the other one involves collecting information about the ecological relationship or correlation between native and invasive species.

Even though traditional methods are widely used, they have some disadvantages. The morphological characters used for the assessment of different species vary greatly between individuals. These methods also fail to identify the species, if the amount of specimen available is too low. Moreover these methods are very time-consuming and require highly trained taxonomists. The lack of availability of trained

taxonomists lengthens the process. Thus, molecular techniques are considered to be the suitable strategy for the identification of invasive species. Variations at molecular level are exploited using these techniques. It also completes the process in quick time, which makes it more preferable than traditional methods. There are a number of molecular techniques available for species identification, for example, stand-alone molecular methods like immunological (Symondson et al. 1999) or protein-based techniques (Soares et al. 2000) and PCR-based molecular diagnostic methods (Dinesh et al. 1993). But the main problems with these techniques are that they are limited to finite range of taxa and the inconsistency in PCR-based techniques. Thus DNA barcoding technique emerged as a powerful tool for species identification and has the potential to overcome all the above limitations (Hebert et al. 2003).

5 DNA Barcoding and Identification of Invasive Species

DNA barcoding stands out of many other molecular tools for expeditious species identification, as it contributes answers to questions that were previously beyond the reach of traditional disciplines. DNA barcode, in its simplest definition, is a short sequence of DNA that in virtue should be easily generated and characterized for almost all species in our planet (Savolainen et al. 2005). A gene region to perform as a DNA barcode should have some salient properties like (1) the typical gene segment should possess notable species-level genetic variability, (2) it should own conserved flanking regions for the generation of universal PCR primers for extended taxonomic applications, and above all (3) it should be a short sequence so that DNA extraction and amplification can be done flawlessly. The COI gene is generally being used for almost all animal species. It has been proved to be highly effective in identifying birds, butterflies, fishes, flies, etc. In contrast, the efficacy of COI gene is not exhibited in case of plants (Kress and Erickson 2007; Eberhardt et al. 2012). Henceforth, the search for effective gene regions for this major group of organism leads to the recognition of a combination of plastid genes rbcL, matK and trnH-psbA.

Biological invasions remain as a biggest peril to the biodiversity after habitat destruction. It put scientists into a tiresome procedure, that is, to define the widely divergent criteria of “invasive species”. Early detection of dawning non-native species enhances the brisk management actions that ensure their eradication (Simberloff 2003; Vander et al. 2010). If the detection of such alien species is pursued after their firm establishment, then the eradication becomes difficult and more expensive (Simberloff et al. 2005). In case of many alien species, the early detection is contradicted by low detectability, which represents the probability that an organism will be observed if present. DNA barcoding is useful for the early detection, where identifiable material is too damaged or present in a negligible amount (Armstrong and Ball 2005; Darling and Blum 2007).

5.1 Pests

Expansion of invasive pests in the geographic range requires quick detection and response because otherwise their management becomes tedious. Hence, DNA barcoding that can lend insights beyond the data obtained through morphological analysis can act as an aid for rapid detection of pest species and thereby enhance the downstream management procedures. Terebrantia and Tubulifera, two recognized suborders of the order Thysanoptera, are terrific pests that are commonly called as thrips. Chilli thrips belonging to the order Thysanoptera, native to South Asia, invaded the United States via commodity shipments and later triggered a huge economic loss to states like Hawai'i (Seal et al. 2010) and Florida (Brown and Osbourne 2008). Also, *Thrips parvispinus*, an insect vector that can transmit tospoviruses, effects to a number of plant species in unrelated plant families used in agriculture, horticulture, etc. On the basis of integrated approach of morphology and DNA barcoding, the invasion of this dreadful pest was reported for the first time in India from papaya plantations (Kaomud et al. 2015). Apart from enhancing the early detection of alien species (Onah et al. 2015), DNA barcoding also helps to find out their source regions (Bellis et al. 2015) as well as their introduction patterns. Mastrangelo et al. (2014) differentiated *Heliothis armigera* (Gram Pod Borer), an invasive pest species from the native *H. zea* (corn earworm) of Brazil using DNA barcoding and thereby disclosed the spread of former species into Brazil. The identification of *Agrilus ribesi* (buprestid beetle), whose invasion to North America had been overlooked for a century, has been detected recently by the application of DNA barcoding (Jendek et al. 2015).

5.2 Aquatic Species

More than two thirds of our Earth is covered with oceans and other water bodies; thus, the evaluation of aquatic species diversity is a demanding task. Also, the increase in global population engenders the tendency of exploiting marine resources for energy, food, etc. Sequentially, this puts thrust on aquatic environment and requests for rapid management. DNA barcoding has been used for the assessment of invasive species that are predominant in the aquatic ecosystem. Aquatic alien species are nonindigenous species that cause peril to the diversity of native species and thereby affect the ecological stability of the infested water body. Floating pennywort (*Hydrocotyle ranunculoides*), member of the Araliaceae plant family, is an invasive aquatic species that is native to North America. This typical invasive species seems to throw serious problems to the waterway management outside of its original allocated area in Western Europe and Australia. Discrimination of *H. ranunculoides* from its related species is a challenging task. Sequencing one of the variable chloroplast loci, namely, trnH-psbA, served to discriminate the invasive

plant species, *H. ranunculoides*, from its closely related congeners like *H. vulgaris*, *H. verticillata*, etc. (Van De Wiel et al. 2009).

Myriophyllum, *Cabomba* and *Ludwigia* which belonged to the family Hydrocharitaceae represent a significant aquatic invasive species. But, many of their related species that are morphologically similar in their vegetative state are non-invasive and are commercially traded. Thus, the detection of invasive plants constituting to this family is a demanding task. The study conducted by Ghahramanzadeh et al. (2013) revealed a non-coding spacer (trnH-psbA) as the best performing DNA barcode for the differentiation of invasive and non-invasive species belonging to Hydrocharitaceae family. The *Dreissena polymorpha* (zebra mussel) and *D. rostriformis bugensis* (quagga mussel) are found to be the most competitive invaders in the freshwaters of North America and Europe. Though the phenotypic plasticity found in the genus *Dreissena* blocks their reliable discrimination as well as identification, the adult *Dreissena* individuals were differentiated with the help of CO1 gene (Jonathan and Karine 2013).

5.3 Plants

The most challenging aspect about invasive plant species is that they scarcely represent a constant, static entity. The knowledge about the introduction of these plant species as well as the actual level of genetic diversity between them is not well explained. Shaik et al. (2016) conducted a study using DNA barcoding, to understand the genetics and ecophysiology of plant invaders of Australia belonging to the family Cucurbitaceae. The *Cucumis myriocarpus* (prickly paddy melon), *Citrullus lanatus* (camel melon) and *Citrullus colocynthis* are very much morphologically related to each other and were found to cause a great infestation to the natural as well as agricultural ecosystem by acting as a harmful weed in several places of Australia. The camel melon and prickly paddy melon are annuals, whereas the *C. colocynthis* is perennial, and hence the management procedures differ among them. The chloroplast gene (ycf6-psbM) and nuclear gene (G3pdh intron region) served as the barcode for interspecific and intraspecific variability among these three cucurbitaceous invasive species (Shaik et al. 2016).

5.4 Mammals

Herpestes auropunctatus, the small Indian mongoose, is a carnivore, which feeds on almost all major vertebrate groups, invertebrates and even plants (Lewis et al. 2011). This typical species which are believed to be native to India and Myanmar (Veron et al. 2007) are often confused with the Javan or small Asian mongoose (*H. javanicus*). All literatures published prior to 2007 presume that the mongoose introduced to Hawaiian and Caribbean islands was *H. javanicus*. Bennett et al. (2011) conducted a study exposing the efficacy of DNA barcoding approaches with mtDNA

cytochrome *b* to discriminate between the two species *H. auropunctatus* and *H. javanicus* as well as the other sympatric members of the genus *Herpestes* (*H. naso*, *H. urva* and *H. edwards*). Further, the study enhanced the confirmation of *H. auropunctatus* invasion in the Hawaiian and Caribbean islands.

6 Conclusion

DNA barcoding has renovated the field of species identification by serving as a robust molecular tool for resolving species diversity. It helps in the accurate identification of specimens, which prevail confusion with traditional techniques. In addition, the availability of open database of DNA barcodes constituting broad ranges of taxonomic groups makes the detection of unknown species easy-going. Barcoding technique that exhibits a significant role in early detection as well as prevention of biological invasions is now been adopted by biosecurity agencies in order to control the spread of invasive species.

Acknowledgements The authors are grateful to Science and Engineering Research Board (SERB), Department of Science and Technology, Government of India, New Delhi, for the financial assistance (EMR/2015/000937).

References

Armstrong KF, Ball SL (2005) DNA barcodes for biosecurity: invasive species identification. Philos Trans R Soc Lond Ser B Biol Sci 360(1462):1813–1823

Bellis GA, Gopurenko D, Cookson B, Postle AC, Halling L, Harris N, Yanase T, Mitchell A (2015) Identification of incursions of Culicoides Latreille species (Diptera: Ceratopogonidae) in Australasia using morphological techniques and DNA barcoding. Austral Entomol 54(3):332–338

Bennett CE, Wilson BS, DeSalle R (2011) DNA barcoding of an invasive mammal species, the Small Indian Mongoose (Herpestes javanicus; E. Geoffroy Saint-Hillaire 1818) in the Caribbean and Hawaiian Islands. Mitochondrial DNA 22(1–2):12–18

Brown SH, Osbourne LS (2008) Chilli Thrips (Scirtothrips dorsalis): a Landscaper's guide. University of Florida, IFAS Extension, Lee County Extension

Carroll SP, Dingle H (1996) The biology of post-invasion events. Biol Conserv 78:207–214

Cock MJW, Kenis M, Wittenburg R (2003) Biosecurity and forests: an introduction-with particular emphasis on forest pests. FAO Forest Health and Biosecurity Working Paper FBS/2E

Darling JA, Blum MJ (2007) DNA-based methods for monitoring invasive species: a review and prospectus. Biol Invasions 9:751–765

Dinesh KR, Lim TM, Chan WK, Phang VPE (1993) RAPD analysis: an efficient method of DNA fingerprinting in fishes. Zool Sci 10:849–854

Eberhardt U, Kress WJ, Erickson DL (2012) Methods for DNA barcoding of fungi. In DNA barcodes: methods and protocols: methods in molecular biology. Humana, Springer, pp 183–206

Ghahramanzadeh R, Esselink G, Kodde LP, Duistermaat H, van Valkenburg JL, Marashi SH, Smulders MJ, van de Wiel CC (2013) Efficient distinction of invasive aquatic plant species from non-invasive related species using DNA barcoding. Mol Ecol Resour 13(1):21–31

Golani D (1993) Trophic adaptations of Red Sea fishes to the eastern Mediterranean environment – review and new data. Isr J Zool 39:391–402
Hebert PDN, Cywinska A, Ball SL, de Waard JR (2003) Biological identifications through DNA barcodes. Proc R Soc Lond 270:313–321
Huey RB, Gilchrist GW, Carlson ML, Berrigan D, Serra L (2000) Rapid evolution of a geographic cline in size in an introduced fly. Science 287(5451):308–309
Jendek E, Grebennikov VV, Bocak L (2015) Undetected for a century: Palaearctic *Agrilus ribesi* Schaefer (Coleoptera: Buprestidae) on currant in North America, with adult morphology, larval biology and DNA barcode. Zootaxa 4034:112–126
Jonathan M, Karine VD (2013) Using DNA barcoding to differentiate invasive Dreissena species (Mollusca, Bivalvia). Zookeys 365:235–244
Kaomud T, Vikas K, Devkant S, Rajasree C (2015) Morphological and DNA barcoding evidence for invasive pest thrips, Thrips parvispinus (Thripidae: Thysanoptera). Newly recorded from India. J Insect Sci 15(1):105
Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. PLoS One 2(6):e508
Lewis DS, Van veen R, Wilson BS (2011) Conservation implications of Small Indian Mongoose (Herpestes auropunctatus) Predation in a Hotspot within a Hotspot: the Hellshire Hills, Jamaica. Biol Invasions 13(1):25–33
Mack RN (1985) Invading plants: their potential contribution to population biology. In: White J (ed) Studies on plant demography. Academic, London, pp 127–142
Mastrangelo T, Paulo DF, Bergamo LW, Morais EGF, Silva M, Bezerra-Silva G, Azeredo-Espin AML (2014) Detection and genetic diversity of a heliothine invader (Lepidoptera: Noctuidae) from North and Northeast of Brazil. J Econ Entomol 107:970–980
McMichael AJ, Bouma MJ (2000) Global changes, invasive species and human health. In: Mooney HA, Hobbs RJ (eds) Invasive species in a changing world. Island Press, Washington, DC, pp 191–210
Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. PLoS Biol 3(12):e422
Meyerson LA, Reaser JK, Chyba CF (2002) A unified definition of biosecurity. Science 295 (5552):44
Mullin BH et al (2000) Invasive plant species. Ames, Iowa: Council for Agricultural Science and Technology. Issue Paper No. 13
Onah IE, Eyo JE, Taylor D (2015) Molecular identification of tephritid fruit flies (Diptera: Tephritidae) infesting sweet oranges in Nsukka agro-ecological zone, Nigeria, based on PCR-RFLP of COI gene and DNA barcoding. Afr Entomol 23(2):342–347
Palumbi SR, Cipriano F (1998) Species identification using genetic tools: the value of nuclear and mitochondrial gene sequences in whale conservation. J Hered 89:459–464
Pimentel D, Zuniga R, Morrison D (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. Ecol Econ 52:273–288
Por FD (1978) Lessepsian migration. Springer, Berlin
Porazinska DL, Pratt PD, Glblin-Davis RM (2007) Consequences of *Melaleuca quinquenervia* invasion on soil nematodes in the Florida Everglades. J Nematol 39:305–312
Ratnasingham S, Hebert PDN (2007) BOLD: the barcode of life data system (www.barcodinglife.org). Mol Ecol Notes 7(3):355–364
Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. Annu Rev Ecol Syst 27:83–109
Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopedia of life: an introduction to DNA barcoding. Philos Trans R Soc Lond Ser B Biol Sci 360 (1462):1805–1811
Seal DRW, Klassen W, Kumar V (2010) Biological parameters of *Scirtothrips dorsalis* (Thysanoptera: Thripidae) on selected hosts. Environ Entomol 39(5):1389–1398

Shaik RS, Zhu X, Clements DR, Weston LA (2016) Understanding invasion history and predicting invasive niches using genetic sequencing technology in Australia: case studies from Cucurbitaceae and Boraginaceae. Conservation physiology, vol 4. Society for Experimental Biology. https://doi.org/10.1093/conphys/cow030

Simberloff D (2003) How much information on population biology is needed to manage introduced species? Conserv Biol 17(1):83–92

Simberloff D, Parker IM, Windle PN (2005) Introduced species policy, management, and future research needs. Front Ecol Environ 3:12–20

Soares RP, Santanna MR, Gontijo NF, Romanha AJ, Diotaiuti L, Pereira MH (2000) Identification of morphologically similar Rhodnius species (Hemiptera: Reduviidae: Triatominae) by electrophoresis of salivary heme proteins. Am J Trop Med Hyg 62(1):157–161

Symondson WOC, Gasull T, Liddell JE (1999) Rapid identification of adult whiteflies in plant consignments using monoclonal antibodies. Ann Appl Biol 34(3):271–276

Van De Wiel CCM, Van Der Schoot J, Van Valkenburg JLCH, Duistermaat H, Smulders MJM (2009) DNA barcoding discriminates the noxious invasive plant species, floating pennywort (*Hydrocotyle ranunculoides* L.f.), from non-invasive relatives. Mol Ecol Resour 9:1086–1091

Vander ZMJ, Hansen GJ, Higgins SN, Kornis MS (2010) A pound of prevention, plus a pound of cure: early detection and eradication of invasive species in the Laurentian Great Lakes. J Great Lakes Res 36(1):199–205

Veron G, Patou ML, Pothet G, Simberloff D, Jennings AP (2007) Systematic status and biogeography of the Javan and Small Indian Mongooses (Herpestidae, Carnivora). Zool Scr 36(1):1–10

Vitousek PM, D'Antonio CM, Loope LL, Westbrooks R (1996) Biological invasions as global environmental change. Am Sci 84(5):468–478

Williams PA, Timmins S (2002) Economic impacts of weeds in New Zealand. Biol Invasions:273–282

Williamson M (1996) Biological invasions. Chapman and Hall, New York, p 244

Part II
DNA Barcoding of Microbes

Microbial DNA Barcoding: Prospects for Discovery and Identification

Anand Mohan, Bableen Flora, Madhuri Girdhar, and S. M. Bhatt

Abstract DNA barcoding is a technology used for the identification of any biological species by sequencing obtained through amplification of DNA. Here we seek a standardized protocol along with choice of markers used for different microbial species: bacteria, fungi, algae, plants and animals. Various projects, consorts, tools and databases have been discussed. With the advancement of technology, the future scope and applications also have been identified with an argument on the limitations along with the concept of universality on the DNA barcoding.

Keywords Microbial · DNA barcoding · Databases · Species identification · Protocol for barcoding

1 Introduction

DNA barcoding is a technology using gene sequences to differentiate species, like the retail stores in which barcodes are used to sell and differentiate the different items. Hebert et al. (2003) state that DNA barcoding is a technique in which species identification is performed by using DNA sequences from small fragment of the genome and resulted in ecological studies in which traditional taxonomic identification is not possible. The short DNA sequences of 400–800 bp long are called DNA barcodes. Hebert and Gregory (2005) stated that DNA barcode is not just any DNA sequence, but it is a standardized sequence of a minimum length and quality from an agreed-upon gene which is deposited in a major sequence database and attached to a voucher specimen whose origins and current status are recorded. The ecological

A. Mohan (✉) · B. Flora
School of Bioengineering and Biosciences, Lovely Professional University, Phagwara, Punjab, India

M. Girdhar
CT Group of Institution, Jalandhar, Punjab, India

S. M. Bhatt
Sant Baba Bhag Singh University, Jalandhar, Punjab, India

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_6

imbalances occur due to global warming leading to loss of diversity of species; therefore, there is a need to identify and classify organisms and study their evolutionary relationships so as to conserve the threatened species. Casiraghi et al. (2010) described that one of the best advantages of a DNA barcode is to associate all life history stages and genders or to identify organisms from part like parasites or to segregate a matrix containing a mixture of biological species. Phylogenetic analysis and searching for the clade along with taxonomy is the chief function of DNA barcoding (Hajibabaei et al. 2007).

Earlier **Carolus Linnaeus** 'father of taxonomy' classifies species that we still use today. In his publication *Systema Naturae*, Linnaeus classified species into hierarchy. He proposed broad groups, called kingdoms to classify microbes, animals and plants. These kingdoms were further divided into phylum, classes, orders, genera (genus is singular) and then species. But as the variations are increasing, it leads to evolution of new species which broaden our biodiversity. So, we need a refined and fast method of classification of the living organism. DNA barcoding brought a revolution in not just classifying the species but also in obtaining many more applications in day-to-day life. It also leads to discovery of new species (Hebert and Gregory Ryan 2005). Raja et al. (2017) demonstrated the utility of DNA barcoding for quality of food by identifying fungi commonly present in dietary supplements. They have studied mushroom samples used in dietary supplements via its fungal barcoding. In DNA barcoding, barcode gap analysis or phylogenetic tree is built. Das and Deb (2015) analysed that a barcode gap exists if the minimum interspecific variation is bigger than the maximum intraspecific variation.

DNA barcoding can be done from small, damaged or industrially processed material, and the DNA barcodes can be efficiently processed and analysed through different databases.

The most relevant tool used for DNA barcoding is **BOLD**, the Barcode of Life Data Systems, comprised of set of integrated databases. It includes Public Data Portal and BIN database acting as primary data source, whereas publication and primer databases support it. One can easily analyse their data in BOLD before submitting it to GenBank, EMBL and DDBJ which are known as the International Nucleotide Sequence Databases. They are the permanent public repository for barcode data records.

2 Origin of Barcoding

Paul Hebert: The Father of DNA Barcoding

Paul D. N. Hebert is a molecular biologist and director of the New Biodiversity Institute of Ontario at University of Guelph in Canada. Hebert is an avid insect collector and admires of biodiversity. Hebert was working on gene sequencing, and most of the people approached him for the sequencing of their soft corals which gave him the idea of barcoding. Governments, museums and universities were focussing to track the invasive species which motivated him. Hebert started his effort at

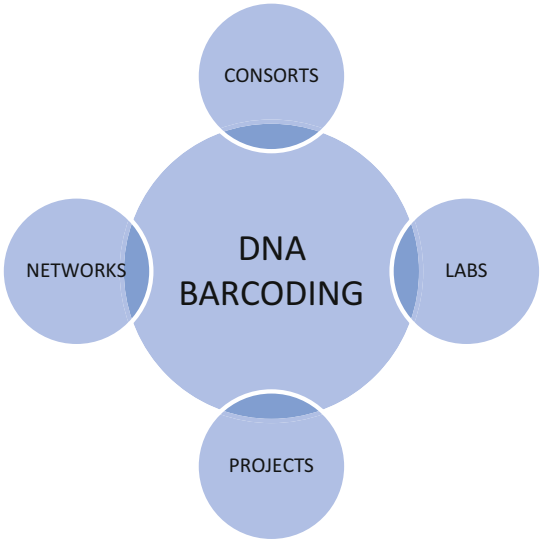


Fig. 1 Diversity of DNA barcoding

Biodiversity Institute of Ontario—the world's first barcode factory—and at present different consortia, databases, labs, networks and projects diversity have been added and showed a successful endeavour as shown in Fig. 1.

2.1 Consorts

2.1.1 iBOL

iBOL—the International Barcode of Life Project. This 25-nation consortium was organized by the Biodiversity Institute of Ontario at the University of Guelph with support from Genome Canada. iBOL's goal is to create five million barcode records from 500,000 species in 5 years. Ten working groups devoted to different taxonomic groups or habitat types form the core of the activity. The International Barcode of Life project (iBOL) is the largest biodiversity genomics initiative ever undertaken. Hundreds of biodiversity scientists, genomics specialists, technologists and ethicists from 25 nations are working together to construct a richly parameterized DNA barcode reference library that will be the foundation for a DNA-based identification system for all multicellular life. In the first step of operations (2010–2015), iBOL collaborators will barcode five million specimens representing 500,000 species. During construction of the barcode library, iBOL participants will also be building the infrastructure needed to use it in real-world situations such as conservation, ecosystem monitoring, forensics and control of agricultural pests and invasive species. Major Consorts have been described and showed in Fig. 2.

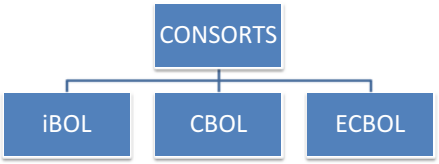


Fig. 2 Major consorts working on DNA barcoding projects

The iBOL global alliance represents a formidable array of nations, research institutes, scientists and private sector partners, reflecting the reality that its research programme demands an alliance between nations with high biodiversity and those with the infrastructure for DNA sequence analysis.

The iBOL global partnership currently involves 25 nations arranged in the following collaborative configuration:

Central Nodes

There are four major central nodes which coordinate iBOL and provide funds. The European Union including France, Germany, Italy, the Netherlands, Portugal, Spain and the United Kingdom and countries such as Canada, China and the United States act as centre for data analysis and its archives.

National Nodes

Argentina, Colombia, Costa Rica, Kenya, Madagascar and Panama will survey national biodiversity for DNA barcoding. They get their species sequencing via central nodes.

Regional Nodes

Countries including Argentina, Australia, Brazil, India, Korea, Mexico, New Zealand, Norway, Russia and South Africa are involved in regional barcodes. Also, New Zealand and Norway will enhance the barcoding programmes in the Antarctic and Arctic, respectively.

2.1.2 CBOL

CBOL—The Consortium for the Barcode of Life is an international supporter of barcoding through various workshops, conferences, outreach activities and working groups. CBOL was proposed by Alfred P. Sloan Foundation in May 2004. CBOL has identified best barcode region for plants, fungi and protist. CBOL has created a social network for barcoding.

2.1.3 ECBOL

ECBOL—The European Consortium for the Barcode of Life functions as a information and coordination hub for the researchers seeking taxonomical sequencing. It was established in 2007 after the Leiden DNA barcoding meeting in Europe and also get funds from EuroBioFund for promoting the identification of large biodiversity in Europe. In addition, ECBOL along with CBOL oranised various meetings in past to boost up the technological advancements in the field of DNA barcoding.

2.2 Networks

There are different networks to promote DNA barcoding. The first national barcode network was launched in Canada and later in different countries like the Netherlands, Mexico and Australia. This networking improves the quality of life and discovery of new species as well as conservation of species.

2.3 Labs

The first largest factory or lab was generated at University of Guelph, CCDB, the Canadian Centre for DNA Barcoding. Moreover CBOL has organized a 'Leading Labs Network', and about 15 labs in 11 countries with technical assistance have been launched.

2.4 Projects

There are many projects working to develop targeted barcode sequence library. CBOL Fungal Working Group, Coral Reef Barcode of Life, FISH-BOL, HealthBOL, MarBOL and MBI are some major projects going on for the barcoding of different species.

3 Protocol

The various steps for DNA barcoding are:

1. Specimen collection—The collection of a specific species from a diversified area is the initial step for the DNA barcoding. An area rich in biodiversity is searched for a particular species need to be sequenced and analysed. Samples can be collected from the field, natural history museums, zoos, botanical gardens, seed banks, etc.

2. Tissue sampling—After collecting the specimen, species is identified and the sample needs to be taken from the particular species, for instance, if fungal species need to be sequenced and the particular colony or cell should be taken for the extraction of DNA.
3. Isolation of DNA—The extraction of DNA for the amplification and sequencing is done in this step. Various methods for extraction and purification of DNA are applied.
4. Polymeric chain reaction (PCR)—The small amount of DNA needs to be amplified with the help of polymeric chain reaction (PCR). In this the extracted DNA is denatured, annealed and renatured. Specific primers are required for the amplification of DNA. The amount of DNA becomes large and can easily be analysed.
5. Analysis and sequencing—A number of copies have been made using PCR for the extracted DNA of specific species. Now the sequencing is done. Different methods used for sequencing are Sanger method and the new-generation sequencing method. The sequence is represented by a series of letters, CATG representing the nucleic acids—cytosine, adenine, thymine and guanine.
6. Managing available data—A sequence is now obtained and we need to create DNA barcodes. Various tools can be used for detecting barcodes. After generating the barcodes, either it needs to get matched with the available data in the barcode library or submit the new data to the databases such as BOLD by creating Excel file.
7. Publishing data—The submitted data is now published and hosted in the various databases which help for further study of a particular species or its conservation.
8. Using barcode data—The various databases act as a barcode library for the various species. This data can be used by different users for their own purposes. There are three types of users present for such data usage. The general public need to read the available data like in institutions for the study of a particular species. The application user usually gets benefit from available data. The application users use the sequence or taxonomic identification for the purpose of conservation of endangered species and maintaining the record of extinct species for ecological balance. This data leads to discovery of new species as the third user, researchers, utilized this data for the study of evolutionary lineages and relationships among different species. The flow chart of steps involved in the DNA barcoding has been described in Fig. 3.

4 Choice for Barcode Marker

It is one of the important parameters for DNA barcoding as different species show different markers best suited for the barcoding. The following are some of the markers used till now arranged as per species:

1. Animals

   In animals, the COI marker is used for DNA sequencing. COI stands for cytochrome C oxidase subunit 1. It is 650 bp long and has single mtDNA.

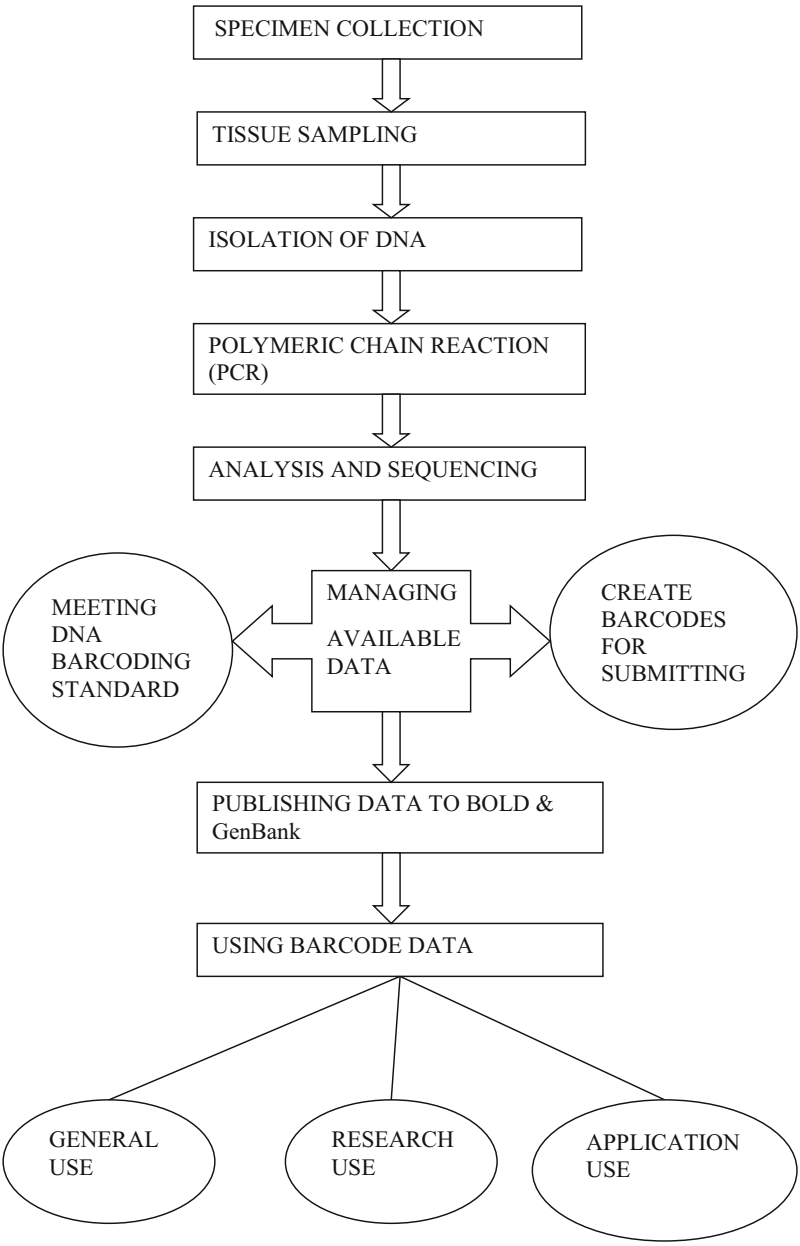


Fig. 3 Flow chart of protocol and application of DNA barcode data

2. Plants

   Plants show low mitochondrial DNA constitution of nucleotides, so COI can't be used in plants. Plastids markers have been used for barcoding in plants (Fazekas et al. 2008).

   Different coding and non-coding regions have standardized the DNA barcoding system. Few coding loci such as rpoB, rpoC1, rbcL, matK and 23S rDNA and non-coding loci such as trnH-psbA, atpF-atpH and psbK-psbI were identified and used as DNA barcoding for plants. Therefore, development of markers for DNA barcoding of plants is still a difficult task as it is hard to find a particular finest combination of marker for a large variety of plant species.

3. Bacteria

   According to the BBC, news scientists have estimated the total number of bacteria on earth to be five million trillion trillion. Such diverse species need to be identified and should be analysed morphologically and phylogenetically. Researchers found the best marker for bacteria is 16sRNA which can be easily amplified and sequenced for making DNA barcodes. QBOL is a consortium of 20 partners including research institutes, universities, etc. from all over the world and shares their research in the field of bacteria, fungi, viruses, nematodes, etc. QBOL is focussing more on the Q-species including genera *Xylella*, *Xanthomonas*, *Clavibacter* and *Ralstonia* affecting the plant health. Since the 1990s, the most common housekeeping genetic marker being used as a marker gene in bacterial identification is 16sRNA. Another gene chaperonin-60 (cpn60) which is 555 bp long was also used as a bacterial barcode marker gene. Chaperonin-60 (cpn60) is also known as GroEL and Hsp60 used for bacterial identification.

4. Viruses

   Viruses are estimated about 10 times more than total number of cells present on earth. Viruses can be sequenced using next-generation sequence technology. In next-generation sequencing, micro- and nanotechnologies are employed to reduce the sample size, lower the cost and enable the parallel sequencing reactions. Arracacha virus, potato black ringspot virus, tomato chocolate virus and others have been sequenced by taking their whole genome. In 2012 watercress and maize lethal necrosis causes severe damage to Kenya maize crop. The watercress white vein has been discovered.

5. Algae

   Several molecular markers were suggested from time to time to identify the algal species. Various markers used as marker for algae are COI for brown and red algae, tufA for green algae, rbcL for diatoms and many more. For carrageenophytes, various molecular markers have been introduced for the molecular taxonomy such as cox1 and cox2–3 spacer, ITS, LSU, rbcL, RuBisCO and the 23S UPA (Conklin et al. 2009).

6. Fungi

Fungus is member of eukaryotic organism that includes microorganisms as yeasts and moulds. Fungi has major role in maintaining the nutrient cycle by breaking down of dead organic material. Fungi can grow with symbiotic manner in plants called mycorrhizae. Moreover many fungi can be used in the formation of drugs, used as food like mushrooms, truffles and morels, and the bubbles in champagne, beer and bread. Fungi also cause various plant and animal diseases. Rusts, smuts and stem and leaf rots are some of the plant diseases causing severe damage to crops. Ringworms and athlete's foot are some of the animal diseases. QBOL focusses on identification of fungi, Viruses, nematodes, phytoplasmas, submitted through DNA barcoding. QBOL has listed 19 Q-species for barcoding. Fungi are the second largest kingdom of eukaryotic life highly diverse and widespread. Fungi help in major production of enzymes, foods and antibiotics industrially for which it contributes economically. There are different kinds of markers analysed and used for the barcoding. Many researchers are trying to find the basic best marker suitable for all the fungi.

Various molecular methods are being used for the phylogenetic analysis and for taxonomical identification. DNA barcoding has provided a better platform for the same but also produced particular challenge for the researcher. Some markers like COI are fit for mushrooms, whereas it is unsuitable for other fungi. So there is a need of a unique marker for the identification of fungi. Many researchers found ITS sequence best for the taxonomical identification for most of the fungi. Fajarningsih (2016) used internal transcribed spacer (ITS) region of nuclear DNA (rDNA) for the identification of various species of fungi. Das and Deb (2015) also described various ITS databases for fungi. ITS region is very much effective for DNA barcoding. They have also analysed various databases for fungi. QBOL, AFTOL, UNITE, MycoBank, etc. are some the major databases for fungi. A newly developed database for fungi has been reported by Irinyi et al. (2015) that focusses on majority of human and animal pathogenic fungi (ISHAM-ITS, freely accessible at http://www.isham.org/ or directly from http://its.mycologylab.org).

Dentinger et al. (2011) has done a comparative study of two different markers COI and ITS and found that COI is not suitable for mushrooms and rusts but may be fit for the other fungi. ISHAM also identified that cox1 is not suitable for DNA barcoding and also described the need of universal marker for the taxonomical advances of fungi. Dulla et al. (2016) ISHAM revealed that majority of medically important species had a low variability in ITS regions. Therefore, ITS sequencing can be utilized for the identification of most medical relevant fungal species. Cox1 found to be unsuitable to be used for DNA barcode (Casiraghi et al. 2010). One of the best advantages of a DNA barcode is to associate all life history stages and genders or to identify organisms from part like parasites or to segregate a matrix containing a mixture of biological species.

5 Tools and Databases for DNA Barcoding

There are various databases and tools used for DNA barcoding. As we require different markers for barcoding, different databases and tools have been created for different species to ease the access. For the large-scale DNA barcoding, funds have been provided by the National Science Foundation (Division of Biological Infrastructure) to develop effective bioinformatics tools. The goal of this project is to develop a standardized bioinformatics tools for large-scale identification of species.

Some of the tools and databases are listed below:

Tool	Organism	Method	Source
AFTOL	Fungi	Structural and biochemical database	https://aftol.umn.edu/
B	Plants	Sequence quality and contig overlap	http://www.nybg.org/files/scientists/dlittle/B.html
BLOG (Barcoding with Logic)	Animals	Diagnostic and character based	http://dmb.iasi.cnr.it/blog-downloads.php
BOLD	Protists, fungi, plants and animals	Taxonomy and identification based	http://www.boldsystems.org/
BRONX	Plants	Sequence identification and taxonomic hierarchy	http://www.nybg.org/files/scientists/dlittle/BRONX.html
CBS prediction servers	All	Sequence analysis	http://www.cbs.dtu.dk/services/
CLOTU	Fungi	Taxa and amplicon data	http://www.bioportal.uio.no
DNA-BAR	Bacteria and fungi	Character-based diagnostic	http://dna.engr.uconn.edu/~software/cgi-bin/barcode/barcode.cgi?num_targets=0
Eco primers	Bacteria, algae	Barcode markers	ecoPrimers.tar.gz
GBIF	All	Open-access database	http://www.gbif.org/
ISHAM	Fungi	Mycological classification	http://www.isham.org/
MEGA software	All	Molecular evolutionary and genetic analysis	http://www.megasoftware.net/
jMOTU		Clustering barcodes	http://www.jmotu.com-about.com/
MycoBank	Fungi	Pairwise sequence alignment and polyphasic identification	http://www.mycobank.org/
OFBG	Plants	Spp. discrimination using oligonucleotide frequencies	http://www.nbri.res.in/ofbg.php
OTUbase	Microbes	Operational taxonomic based	http://www.bioconductor.org/packages/release/bioc/html/OTUbase.html
P.A.T.H.S	Fungi, plants	Species relationship	http://barcoding-paths.it

(continued)

Tool	Organism	Method	Source
Q-Bank	Plant pest	Identification and disease detection database	http://www.q-bank.eu/
TaxI	Plant	Distance based	axel.meyer@uni-konstanz.de
Taxonerator	Microbes	Taxonomy based	http://www.taxonnerator.com-about.com/
UNITE	Ectomycorrhizal fungi	Species hypothesis	http://unite.ut.ee

6 Significance of DNA Barcoding

A DNA barcoding has brought a vital revolution in the world of taxonomy of species. It has contributed in three ways as illustrated by Casiraghi et al. (2010):

1. Molecularization—It involves the use of the variable molecular marker acting as a discriminator for various species. This segregates the species at molecular level.
2. Computerization—Various computer-based databases used to keep the record of various species as well for the phylogenetic analysis.
3. Standardization—Various standards have been made to approach variability among the species.

7 Limitations

1. Lack of Universality
   DNA barcoding also gave rise to several challenges and a collaborative effort is being made for the amendments in the protocol. There is a need of multiple analyses; Blaxter (2004) found the necessity of core plurality with more than one sequence in one sample. This will improve DNA preservation and taxonomic units. There is no universal gene for DNA barcoding, no single gene that is conserved in all domains of life and exhibits enough sequence divergence for species discrimination (Purty and Chatterjee 2016). New species have been developed by hybridization, and their taxonomical and molecular analyses are required which can be possible, but a single marker is required for the detection of their variations (Ali et al. 2014).
2. Co-amplification
   There is a need of combined amplification to fasten the process as well as for better analyses of the sequences.
3. Lack of Taxonomic Experts
   Taxonomic experts are required for the correct analysis of the submitted data as well as for the discovery of new species which leads to ecologically and economically advancements.

8 Applications and Future Aspects

1. Controlling the Pest
   DNA barcoding helps in the control of various pests by identifying its taxonomy and prevents the huge amount of farm production. Fruit flies have been controlled by global tephritid barcoding identification.
2. Identification of Disease Vectors
   DNA barcoding helps to identify the vector species that causes infectious diseases to animals and humans. This leads to easy detection of diseases thus can be cured with effective way. A global mosquito barcoding is initiated for the building of a reference barcode library that can help to diagnose and cure the chronic diseases. Abel et al. (2015) and their colleagues used STAMP to trace the growth and decline of *V. cholerae* infection in rabbits. They have added about 500 different barcodes for the batch of *V. cholerae*. Animal products legally used in production of various medicines can be analysed and separated from those used illegally. Yan et al. (2013) developed a DNA barcoding technique that can be used to discriminate between a wide range of animal species whose horns are used in traditional medicines.
3. Conservation of Natural Resources and Biodiversity
   Natural resources need to be conserved. The illegal products from hardwood tress for the trade purpose and illegal trade of endangered species should be stopped. Using DNA barcoding, the endangered species can be easily identified and conserved. Databases like FISH-BOL act as reference library to improve the management and conservation of natural resources. About 90% primate population has been reduced due to hunting for bush meat. DNA barcoding helps to prevent the endangered species. Devloo-Delvaa et al. (2016) identified a biological pollutant, pygmy mussel (*Xenostrobus securis*), and nearly 130 specimens are analysed to obtain the genetic variability and new sources of pollution. Soil contains huge amount of biodiversity. DNA barcoding helps to identify various species and use them in a right way. Various national and international projects have been established to understand the biodiversity of soil focussed by Orgiazzi et al. (2015).
4. Enhanced Quality of Water
   DNA barcoding helps in taxonomical advances of organism present in rivers, streams, lakes and ponds. It enhances quality of water by detecting the species obtained and helps in cure of water-borne diseases.
5. Improvement of Food Quality
   The packed food quality can be improved by deep study of a particular species. Trivedi et al. (2016) assessed cryptic species in marine environment. Various projects on identification of marine organism are MarBOL, CeDAMar, CMarZ, SHARK-BOL, etc. This helps to find the various food-borne species and their effects on the quality of food. Many researchers are trying to find out various edible species of fungi and track the adulterations in the food. A protein for body building and antibiotics contain large number of products from exploitation of

animals. Staats et al. (2016) centralized the concept of metabarcoding for tracking various adulterations and analysing samples from alleged wildlife crime incidents.

6. Generation of New Tools and Software

   Scientists are moving at a faster rate to achieve the identification of variety of species and to detect cryptic species to conserve the biodiversity. New databases and software have been discovered for the faster access and maintenance of large amount of data. Generation of new tools and software will enhance the process as well as create a new source of income as a future scope.

9 Conclusion

After so many years of DNA barcoding, the approach is still controversial. The lack of universality makes it a difficult task to be resolved soon. A single marker can't be used for all the species as different species show different levels of amplification for the different region. In plants chloroplast DNA is used, whereas in animals mitochondrial DNA is used. In the case of microbial barcoding, different species show different effects on the choice of markers. For instance, bacterial barcoding can be done using marker 16sRNA, whereas COI is best for algae. Even the classes in a particular species also show the need of a single marker as the classes also have different markers like in the case of fungi. A basic protocol of extracting DNA, sequencing and submitting or analysing of DNA barcoding has been discussed.

Success of the DNA barcoding can be estimated from the increasing amount of barcoding projects and consorts which include a large number of scientists from all over the world. Various applications in food security and identification of pathogen lead to the easy diagnosis of various diseases. The phylogenetic relationship and the clade information can be analysed using DNA barcoding.

The BOLD data system is the main axis to the barcoding. BOLD assembles standard information on voucher specimen and its current status which allows advancements in the findings of deposited sequences.

Bioinformatics has contributed at a large scale. Different software and tools have been developed for DNA barcoding to ease the process. This leads to easy comparison between genomes with high-throughput technology as well as efficient system which includes large-scale assay of species. This improves the research which will further have great impact in the field of medicine, agriculture, ecology development and biodefence.

References

Abel S, Abel zur Wiesch P, Chang H-H, Davis BM, Lipsitch M, Waldor MK (2015) Sequence tag–based analysis of microbial population dynamics. Nat Methods 12:223–226. https://doi.org/10.1038/nmeth.3253

Ajmal Ali M, Gyulai G, Hidvegi N, Kerti B, Al Hemaid MAF, Pandey KA, Lee J (2014) The changing epitome of species identification – DNA barcoding. Saudi J Biol Sci 21(3):204–231. https://doi.org/10.1016/j.sjbs.2014.03.003

Blaxter ML (2004) The promise of a DNA taxonomy. Philos Trans R Soc Lond B 359:669–679

Casiraghi M, Labra M, Ferri E, Galimberti A, Mattia FD (2010) In: Nimis PL, Vignes Lebbe R (eds) Tools for identifying biodiversity: progress and problems, pp 269–73

Conklin KY, Kurihara A, Sherwood AR (2009) A molecular method for identification of the morphologically plastic invasive algal genera Eucheuma and Kappaphycus (Rhodophyta, Gigartinales) in Hawaii. J Appl Phycol 21:691

Das S, Deb B (2015) DNA barcoding of fungi using ribosomal ITS Marker for genetic diversity analysis: a review. Int J Pure Appl Biosci 3(3):160–167

Devloo-Delvaa F, Mirallesa L, Ardura A, Borrell YJ, Pejovic I, Tsartsianidou V, Garcia-Vazquez E (2016) Detection and characterisation of the biopollutant Xenostrobus securis (Lamarck 1819) Asturian population from DNA barcoding and eBarcoding. Mar Pollut Bull 105(1):23–29

Dentinger BTM, Didukh MY, Moncalvo J-M (2011) Comparing COI and ITS as DNA barcode markers for mushrooms and allies (Agaricomycotina). PLoS One 6(9):e25081. https://doi.org/10.1371/journal.pone.0025081

Dulla EL, Kathera C, Gurijala HK, Mallakuntla TR, Srinivasan P, Prasad V, Mopati RD, Jasti PK (2016) Highlights of DNA barcoding in identification of salient microorganisms like fungi. J Mycol Med 26(4):291–297. https://doi.org/10.1016/j.mycmed.2016.05.006

Fajarningsih DN (2016) Internal Transcribed Spacer (ITS) as DNA barcoding to identify fungal species: a review. Squalen Bull Mar Fish Postharvest Biotech 11(2):37–44

Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG et al (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS One 3(7):e2802

Hajibabaei M, Singer CAG, Hebert NDP, Hickey AD (2007) DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends Genet 23(4):167–172

Hebert DNP, Gregory RT (2005) The promise of DNA barcoding for taxonomy. Syst Biol 54(5): 852–859

Hebert NDP, Cywinske A, Ball LS, dewaard RJ (2003) Biological identifications through DNA barcodes. Proc R Soc Lond B 270:313–321

Irinyi L, Lackner M, de Hoog S, Meyer W (2015) DNA barcoding of fungi causing infections in humans and animals. Fungal Biol 120:125–136. https://doi.org/10.1016/j.funbio.2015.04.007

Kress WJ and Erickson DL (2007) DNA barcodes: methods and protocols. Methods Mol Biol 858 (chapter)

Orgiazzi A, Dunbar BM, Panagos P, Groot DAG, Lemanceau P (2015) Soil biodiversity and DNA barcodes: opportunities and challenges. Soil Biol Biochem 80:244–250

Purty RS, Chatterjee S (2016) DNA barcoding: an effective technique in molecular taxonomy. Austin J Biotechnol Bioeng 3(1):1059

Raja AH, Baker RT, Little GJ, Oberlies HN (2017) DNA barcoding for identification of consumer-relevant mushrooms: a partial solution for product certification? Food Chem 214:383–392

Staats M, Arulandhy JA, Gravendeel B, Holst-Jensen A, Scholtens I, Peelen T, Prins WT, Kok E (2016) Advances in DNA metabarcoding for food and wildlife forensic species identification. Anal Bioanal Chem 408:4615–4630

Trivedi S, Aloufi AA, Rehman H, Saggu S, Ghosh SK (2016) DNA barcoding: tool for assessing species identification in reptilia. J Entomol Zool Stud 4(1):332–337

Yan D, Luo JY, Han YM, Peng C, Dong XP et al (2013) Forensic DNA barcoding and bio-response studies of animal horn products used in traditional medicine. PLoS One 8(2):e55854. https://doi.org/10.1371/journal.pone.0055854

DNA Barcoding on Bacteria and Its Application in Infection Management

Mohammad Zubair, Farha Fatima, Shamina Begum, and Zahid Hameed Siddiqui

Abstract The development of bacterial DNA barcoding with error-independent and time-saving technique has increased dramatically, which helps in identifying the microbes, understanding the microbial biodiversity, and analyzing the diseases, related to these pathogens. The study has discussed the developmental stages of DNA barcode research and compared it with the gold standard methods. It has evaluated the possible application of magnetic- and nanoparticle-related applications in bacterial DNA barcoding. These innovative techniques help in the identification of a number of infectious pathogens, which are present in trace amount. Bio-imaging detection of an infectious microorganism has proved to be effective for the development of fluorescent nanoparticle, super-paramagnetic nanoparticle, and metallic nanoparticle. The living organisms, present with functional materials, have a vast application in clinical research of water-soluble conjugated polymers. However, the performance of DNA barcoding in laboratory scale is easy, and it tends to give maximum benefit to the community, if they developed a global system to share their findings and interpret the sequenced data by genome.

Keywords Bacteria · DNA barcoding · Infection management

Abbreviations

Bg	*Bacillus globigii*
BLAST	Basic local alignment search tool

M. Zubair (✉) · S. Begum
Department of Medical Microbiology, Faculty of Medicine, University of Tabuk, Tabuk, Kingdom of Saudi Arabia

F. Fatima
Department of Zoology, Faculty of Life Science, Aligarh Muslim University, Aligarh, India

Z. H. Siddiqui
Department of Biology, Faculty of Science, University of Tabuk, Tabuk, Kingdom of Saudi Arabia

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_7

BOLD	The barcode of life data system
CPNs	Conjugated polymer nanoparticles
CTX	Cefotaximases
E. coli	*Escherichia coli*
FRET	Fluorescence resonance energy transfer
kDa	Kilo daltons
cpn60	Chaperonin 60
MLST	Multilocus sequence typing
MOTU	Molecular operational taxonomic unit
MTB	*Mycobacterium tuberculosis*
NCBI	National Centre for Biotechnology Information
mecA	Novel penicillin-binding protein, PBP-2a
NPs	Nanoparticles
NTM	Nontuberculous mycobacteria
PAIs	Potential pathogenicity islands
PCR	Polymerase chain reaction
QBOL	The quarantine barcoding of life
RIF	Replication initiation factor
S. aureus	*Staphylococcus aureus*
SHV	Sulfhydryl variable enzymes
Tuf	Elongation factor Tu

1 Introduction

Recently, the extraction of genomic information from the samples of patients is considered as a major development among researchers and healthcare professionals. The development in research areas is emerged from the advancement in ultra-high-throughput sequencing (UHTS) technologies. However, this genomic analysis gives a boost in the scientific and diagnostic research, when the genome size was particularly small. The impressive increase in the study of genomic data and molecular research in clinical microbiology has been evolved as an important tool in diagnostic laboratories (Figs. 1 and 2).

Short DNA sequences are sequentially used for identifying species level as it is emerged as an accurate, standardized, and fast method. In 2003, Hebert et al. (2003) have proposed a new DNA barcoding technique (Hebert et al. 2003). The core intention of this study is the species identification based on a short segment of DNA, extracted from a standardized area of the specified genome. The validation of DNA sequence is practically observed identifying different species. A major advantage of DNA barcoding is the ability to enrich the biodiversity studies, such as discovery and species identification. The relationships between the genetic and evolutionary aspects are studied effectively by researchers using the DNA barcoding. These relationships are extracted from cross-reading distributional, molecular, and

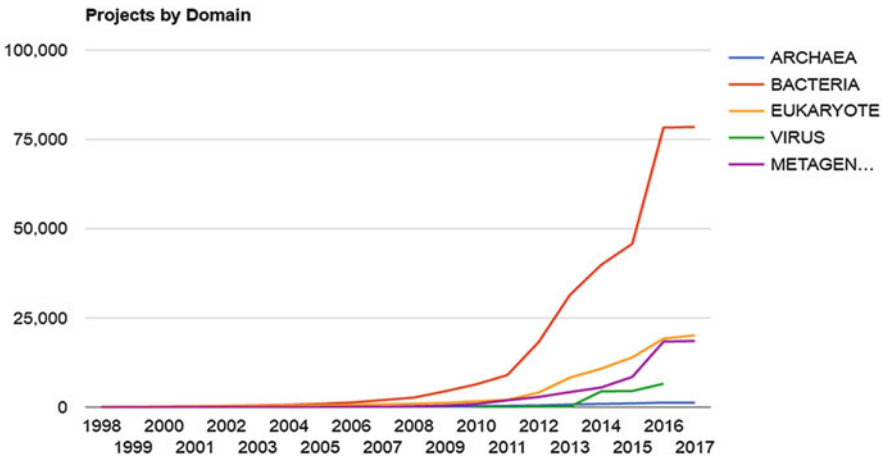


Fig. 1 Milestones in whole-genome sequencing. The number of complete and draft genomes from the archaea ($n = 1285$, year = 2016, blue color), bacteria ($n = 18{,}236$, year = 2012, red color), eukaryote ($n = 13967$, year = 2015, yellow color), virus ($n = 6575$, year = 2016, green color), and metagenome ($n = 8492$, year = 2015, violet color), deposited in public databases from 1997 to 2017, is shown, as extracted from the GOLD database (www.genomesonline.org/) in January 2017

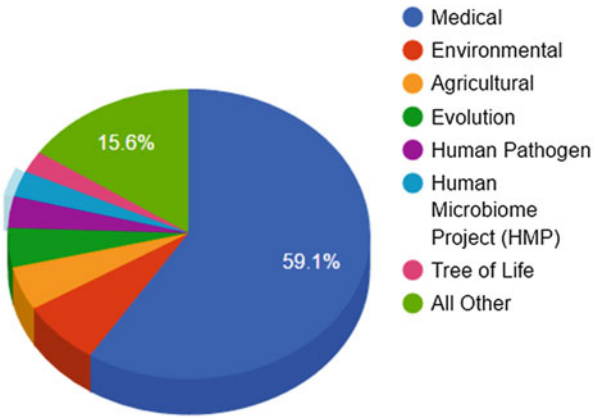


Fig. 2 Distribution of bacterial DNA barcoding project. The percentage of bacterial projects according to their subject in the following order: medicine 59.1% ($n = 31{,}701$), environment 7.2% ($n = 3851$), agriculture 4.9% ($n = 2629$), evolution 4.4% ($n = 2364$), human pathogen 3.5% ($n = 1881$), Human Microbiome Project 2.9% ($n = 1562$), Tree of Life 2.4% ($n = 1304$), and all other remaining 15.6% ($n = 8350$) in 2017 as extracted from the GOLD database (www.genomesonline.org/) in January 2017

morphological data (Savolainen et al. 2005). The identification by DNA barcoding at species level is performed by using a short DNA sequence from a standard part of the genome (Hebert et al. 2003). Hajibabaei et al. (2007) have compared a library of reference barcode sequences with barcode sequence extracted from specified identity

individuals obtained and unidentified specimen, respectively. In contrast, standardized identification method implies an adverse impact on the species mapping on the entire earth. The emergence of particular adverse impact is shown when DNA sequencing technology is economically extractable. DNA barcode recommends that taxa are identified as an 11-digit universal product code by the standardized DNA sequences, identifying market's retail products (Neigel et al. 2007).

The possession analysis, publication, and storage of DNA barcode records are assisted from a database of informatics worktop: the Barcode of Life Data System (BOLD). The association of BOLD is particularly seen with traditional bioinformatics by obtaining distributional, morphological, and molecular data. The researchers having firm knowledge about DNA barcoding can accessibly use BOLD. Ratnasingham and Hebert (2007) indicate that affording specialized services are used linked to BOLD, helping the researchers to acquire barcode designation in the sequence databases internationally. BOLD serve as the universal starting point, which is ultimately referred to specialized databases for the identification of species, for example, pathogenic strains, endangered species, and disease vector species (Ball and Armstrong 2006). BOLD assist DNA barcoding study, which is an online resource available to the scientific community.

The DNA barcode date of bacteria was acquired by QBOL (quarantine barcoding of life) to quarantine. Species quantification occurs by using total DNA barcode, which determines the composition of an insect's bacterial symbiotic association and how they alter in time. It also inspects novel bacterial pathogens of insect pests and assesses the hidden biodiversity of soil samples. More than 170 member organizations are included from 50 countries after the launching of the Consortium for the Barcode of Life (CBOL) in 2004. DNA barcoding sensu stricto is promoted by CBOL as a global standard to identify the biological specimens (Miller 2005). In clinical bacteriology, it is very important to identify the pathogen present in a clinical sample rapidly. This helps in improving patient care reducing the economic burden. Moreover, it is essential to recognize the presence of a pathogen in a clinical sample, specifically in clinical bacteriology. The importance of testing is revealed from the management of infections and antibiotic treatment for species-level identification and antibiotic susceptibility performance.

A major uprising has been identified in clinical microbiology, the matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (MS), implying positively in identification, typing, and toxin detection (Croxatto et al. 2012). The lacuna in this method rapidly approaches to characterization of any bacterial strain. In this regard, multiple analyses were involved under such types of explicit categorizations, and therefore, research laboratories execute a given pathogen specifically. Depending on bacterium type and their complexity of the question, it takes days to months for identifying the bacterium. The chances of minimizing the extent of steps required the pathogen's full characterization and the "time to result" optimization (Fig. 3). The bacterial genomic data of the whole sequence can be obtained from a pure culture, directly from a sample of clinical origin (pus, sputum, body fluids, urine, stool, tissue, blood, serum, etc.) or from a single bacterium present

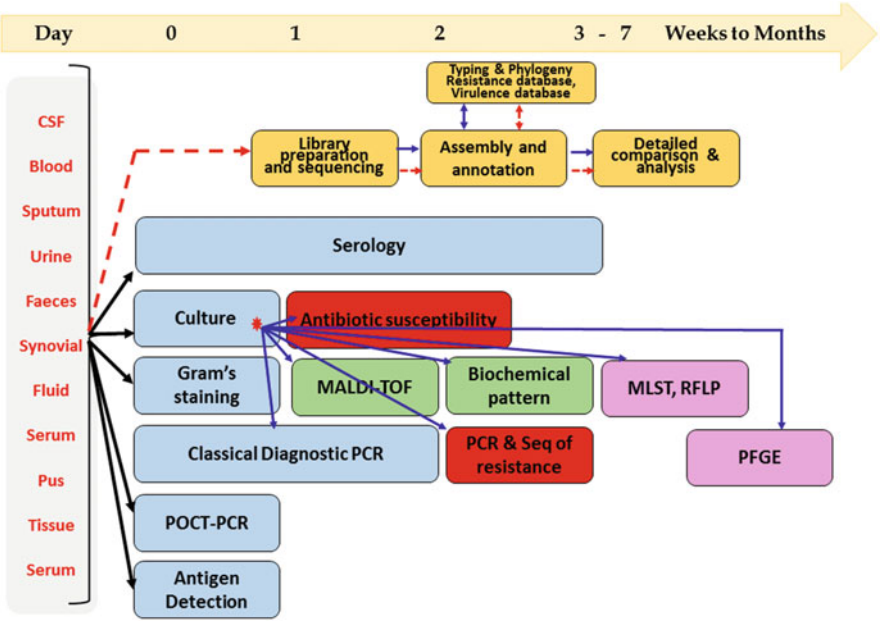


Fig. 3 Schematic representation of the timeline for the processing of clinical samples with classical pathogens. Clinical samples are directly submitted for antigen detection (immune chromatography), PCRs, Gram stain, and/or serology while the bacterium is being cultured. Once a pure bacterial culture is obtained (red star), the second panel of analyses is performed. Direct sequencing allows shortening of the "time to report result" (red dashed arrow), especially for fastidious bacteria or slow-growing bacteria. Techniques are colored according to their application for bacterial detection (blue), species identification (green), antibiotic susceptibility testing (red), and strain typing (pink). The main steps of genome sequencing are highlighted in purple. MALDI-TOF MS, matrix-assisted laser desorption ionization time-of-flight mass spectrometry; MLST, multilocus sequence typing; PFGE, pulsed-field gel electrophoresis; POCT, point-of-care tests; RFLP, restriction fragment length polymorphism

in a clinical sample. In addition, the most costly affair is the genome finishing as it consumes a major extent of time in each phase of a clinical process.

2 Great Criticism on DNA Barcoding

DNA barcoding has been excessively argued by people after its launching. The perception of these people reveals that the traditional methods and taxonomists might vanish through a DNA barcode species identification. In contrast, high-throughput facilities vacuum funding for affording barcode species (Rubinoff et al. 2006). Even though establishment of DNA barcoding is determined as a global bio-identification and its assumption is still reasonably argumentative (Hickerson et al. 2006), the barcode data is emerged as phylogeographic (Beheregaray 2008).

3 DNA Barcoding of Bacteria

Pathophysiology give a clear picture of how bacterial population were grouped in the World based on pathogenesis, neglecting the bacteria to bacteria signalling mechanism. There is a paucity of investigating the minimal extent, initiating the infection in a specific organ region. Thereby, the differentiation of clones in a mixed bacterial population should be assessed by developing a marker throughout a prolonged infection. The specific offerings of tagged mutagenesis were assured for the bacteria pools' nature and preferably involved individual mutants through transposon insertions. According to Mecsas et al. (2001), a unique oligonucleotide barcode is identified in each insertion, which offers the sequential path for individual mutants' fate. This method has been used widely for studying virulence factors that are essential for the identification of numerous bacterial pathogens with their stages of infection. This specific approach has allowed the researchers to examine the individual bacteria's fate during infection.

Phytoplasmas of bacteria cause a huge loss to agriculture production. The universal DNA barcode-based is utilized as an elongation factor Tu (tuf) gene for phytoplasma identification. New primers amplify 420–444 bp fragment of tuf from all 91 phytoplasmas tested strain (16S rRNA groups -I through -VII, -IX through -XII, -XV, and -XX) (Makarova et al. 2012). *Xanthomonas*, a genus with many important phytopathogens, differentiate bacteria at the species level or below, which is initially associated with plants by using DNA markers. These markers have portions of DNA replication initiation factor (RIF). DNA is a single-copy gene consisting of huge majority of sequences of bacterial genome and gene-specific primers needed for RIF amplification. The RIF marker system should be expandable to most bacterial genera, including *Pseudomonas* and *Xylella* (Schneider et al. 2011a).

Ecology of the microbial community is an area of the fast-growing discipline of microbiology. Color-coded enterobacterial taxa are used along with broad-host-range plasmids for creating an artificial microbial community in the laboratory. Moreover, green fluorescent protein color variants are encoded and microbial fitness are developed by these plasmids. The possibilities have been created for community members to respond differently to the external events by understanding the community composition. In addition, the fitness can be measured by understanding the use of community members' responses. The results obtained by pre- and posttest helped to understand about the microbial communities. Barcoding of selected enterobacterial species using fluorescent proteins provides a simple and speedy method for distinguishing species identity (Le et al. 2013).

4 Applications of DNA Barcoding in Clinical Microbiology

In clinical bacteriology, the process to study the clinical samples has changed minutely over the years. However, majority of the analysis still depend on their gold standard methods known as isolation of a viable microorganism (Figs. 3 and 4).

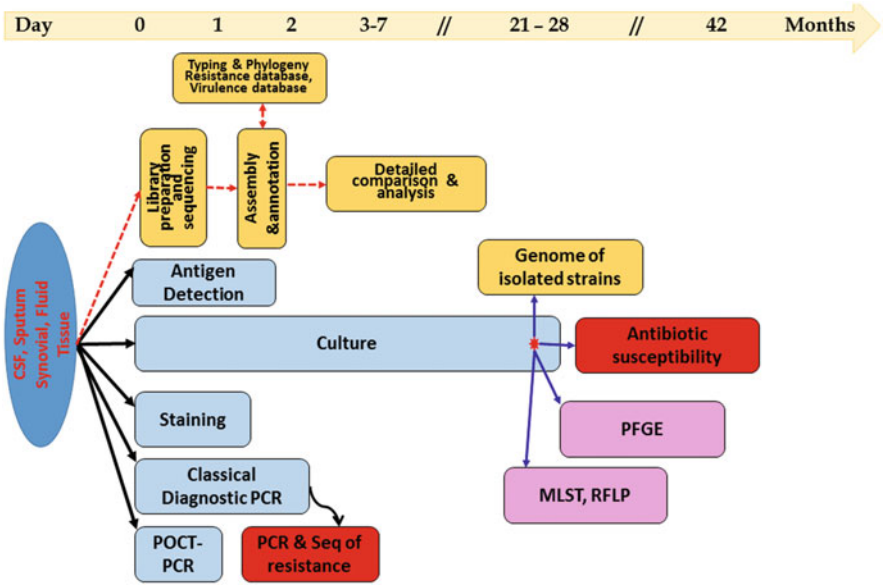


Fig. 4 Schematic representation of the timeline for the processing of slow-growing bacterial samples such as *Mycobacterium tuberculosis*. Clinical samples are directly submitted for antigen detection (immune chromatography), PCRs, Gram stain, and/or serology while the bacterium is being cultured. Once a pure bacterial culture is obtained (red star), the second panel of analyses is performed. Direct sequencing on clinical sample allows shortening of the "time to result" (red dashed arrow), especially for fastidious bacteria or slow-growing bacteria. Techniques are colored according to their application for bacterial detection (blue), species identification (green), antibiotic susceptibility testing (red), and strain typing (pink). The main steps of genome sequencing are highlighted in purple. MLST, multilocus sequence typing; PFGE, pulsed-field gel electrophoresis; POCT, point-of-care tests; RFLP, restriction fragment length polymorphism

The PCR detection method has developed strongly in recent years. The effectiveness of PCR detection method has allowed researchers to ensure identification of viruses, fastidious organisms, intracellular bacteria, and faster diagnosis. The division of genomics' applications is classified into two types. The determination of biological properties, pure bacterial isolate, and outbreak monitoring is required in type 1, while clinical sample, including community profiling and metagenomics, is directly applied in type 2.

The future of clinical microbiology depends on the advancement for the development of methods, which will be helpful in obtaining full genomic data of each and every microbe preferable from clinical samples. The sequence of *Plasmodium falciparum* was done from the cell-depleted samples (Auburn et al. 2011) and the bioterrorism agent *Francisella tularensis* from abscess pus (Kuroda et al. 2012). Samples with high bacterial concentrations or physiological sterile are identified by direct sequencing. If the bacterial concentration in the clinical sample was less, it will be increased by using an initial antibody-based bacterial purification step. Major

advantages are provided from the direct clinical sample sequencing, which offer additional time comparatively to the time required for bacterial culture. It also allows the detection of unculturable bacteria, which are very difficult to culture and require longer time to culture.

The significant contributions to assorted regions in virology are enabled by high-throughput sequencing. The specific areas include pathogenesis, molecular epidemiology, virus discovery, host immune system, antiviral pressures, and metagenomics. According to Quinones-Mateu et al. (2014), hepatitis C virus (HCV), influenza virus, and human immunodeficiency virus (HIV) are studied excessively in the DNA barcoding studies. During the last 15 years, the barcoding revealed a new series of enhancements for fungal research in the form of molecular phylogeny. Moreover, a positive consequence on ecology and fungal biodiversity research has been emerged sequentially through the development of these methods. The progression of acquiring high-quality sequence databases is the reliance factor for fungi identification and technology-driven usability of DNA barcoding. It is examined that taxonomists contribute effectively in curating these high-quality sequence databases (Begerow et al. 2010).

4.1 Clinical Diagnostics and Species Identification

One of the most crucial step in developing accurate clinical decisions is the identification of species. The effectiveness of this approach reveals from the provision of direct information of pathogens strength. Gram staining, biochemical tests, sugar assimilation/fermentation, and colony growth time and morphology are the reliance factors of the gold standard method for identifying bacteria. MALDI-TOF MS is a very low-cost method introduced successfully for routine use (Croxatto et al. 2012). In contrast, approximately unusual species in 50% cases were undetected by MALDI-TOF (Bizzini et al. 2011). In addition, the computational methods of DNA barcoding were proposed to identify these species, explaining the bacterial/archeal 16S barcode loci benchmark. These loci can be utilized to compare the performance of existing and new methods. The benchmark results indicated that the taxon coverage of reference sequences, which were used as a reference, is incomplete for identifying up to genus or species level. The registration of reference barcode sequences needs to be stepped up as it would help in the identification of species level and high-throughput DNA barcoding research (Tanabe and Toju 2013).

MOTU (molecular operational taxonomic unit) is considered as fastest and cost-effective identification method. It is helpful for those that are not encountered before in analyzing the evolutionary patter. MOTU approach possesses problems, particularly in the area of choosing what level of molecular difference defines a biologically relevant taxon (Blaxter 2004). According to Balasingham et al. (2009), a new diagnostic test for the known pathogens has been developed by genome sequences. Moreover, mutant variants such as *Chlamydia trachomatis* harbor a deletion on its

cryptic plasmid (Fenollar et al. 2007) or emerging pathogens such as *Tropheryma whipplei* (Ripa and Nilsson 2007). Raw genome sequence may be helpful to identify the new targets for diagnostic PCR or for the development of the serological test (Greub et al. 2009).

Both genetic and phenotypic attributes have defined the species and fundamentally illustrated by type strains consumed in two international strain centers. It has been proposed that a reference standard should be constituted by reference sequences (Didelot et al. 2012). By using phylogenetic analysis, the classification of the newly sequenced microorganisms should be identified among sequenced reference genomes. The available 1900 bacterial genomes are classified by using multilocus sequence typing (MLST) based on ribosomal protein-encoding genes (Jolley et al. 2012). The presence of bacterial essential genes is considered as an immense advantage in all bacteria, offering additional resolution as compared to 16S rRNA and strongly corroborated for functional preservation. The current classification could cover newly identified genome by investigating the 53 conserved ribosomal genes. Larsen et al. (2012) developed a web-based method for MLST on preassembled genomes or directly on short sequence reads. Currently, only 66 bacterial species have been identified based on the various alleles by MLST schemes, but it covers the most important clinical pathogens.

4.2 Monitoring of Pathogens During an Outbreak

The outbreak of the virulent *E. coli* O104:H4 is an excellent example of the speed of data generation and implementation of controlled measures obtained by whole-genome sequence (WGS). Another outbreak in Germany of diarrhea and hemolytic-uremic syndrome (HUS) is caused by Shiga toxin produced by *E. coli* strain in May 2011. Within 2 months, more than 3100 non-HUS cases were reported, causing 53 deaths. Many independent research centers rushed to sequence some outbreak and historical reference isolates to combine different sequencing technologies (Mellmann et al. 2011; Rasko et al. 2011; Rhode et al. 2012). The first genome was sequenced and deposited in public databases with <2 weeks from the identified onset of the outbreak and in <62 h from strain isolation (Mellmann et al. 2011). The execution of an open-source is considered as an interesting option for genomics program and to identify the strain (Rhode et al. 2012). This occurrence was due to the unusual *E. coli* serotype O104:H4, comprising genes from both enterohemorrhagic *E. coli* (EHEC) and enteroaggregative *E. coli* (EAEC) (Mellmann et al. 2011; Rasko et al. 2011; Rhode et al. 2012). The gain and loss of plasmid-encoded and chromosomal factors can emerge a highly pathogenic hybrid of EHEC and EAEC (Mellmann et al. 2011). The formation of new virulent pathogens occurred from the plasticity of bacterial genomes, respectively, to genetic exchanges. The usefulness and feasibility of rapid draft genome sequencing are demonstrated by this multicenter work.

4.3 Epidemiology Study

Research on epidemiology dynamics of microbes is performed by using two genes (16S rRNA and 60 kDa chaperonin proteins cpn60) of gene-centric metagenomics study to characterize and identify the unknown bacteria. The outbreak efforts of the International Barcode of Life allow to evaluate the protein-coding cpn60 gene and bacteria from the 16S rRNA gene through DNA barcodes (Links et al. 2012). In their findings, this cpn60 universal target has larger barcode gap as compared to 16S rRNA, inferring that cpn60 has been used as preferred barcode for bacteria. A cohesive target for identifying the characterization of species-level data is revealed through this large gap. According to Links et al. (2012), DNA barcode is an authentic technique for evaluating novel microbes in epidemiological studies.

The DNA barcode method facilitates the entire community phylogenetic study of bacteria by using a primer, which was available universally (Barberan and Casamayor 2010; Wang et al. 2012). Small subunit of 16s rRNA was commonly used as gene barcode for the study of bacterial and archaeal communities. The majority of the research used this marker as a barcode to quantify the microbial load in the community from the known environmental samples based on DNA sequences (Hugenholtz et al. 1998). We have only studied the scrape of the microbial community where millions of bacteria and billions of bacterial species exist (Curtis et al. 2002). Only a little part is sequenced scientifically and submitted in the sequence databases (Wu et al. 2009).

4.4 Virulence Study

The bacterial virulence factor provides information that is important for sequence methods. Opening an information bank for large-scale research based on genomic-phenotypic associations helps to predict the resistance to specific antibiotic or general antibiotic susceptibility pattern of particular genera. For instance, mecA confers methicillin resistance to *S. aureus* (Wolk et al. 2009), CTX-M, TEM, and SHV by *Enterobacteriaceae* (Zubair et al. 2012) or drug targets such as rpoB and others for *Mycobacterium tuberculosis* (Hillemann et al. 2005). It is examined that if the entire genome is accessible for investigating the mechanism of resistance, the susceptibility is revealed as a fast and an economical for rapid-growing bacteria. Moreover, gene mutations are not associated directly with the rapid detection of resistance. The bacterial resistance is a complex matter, and this was not studied extensively by barcoding methods (Iacono et al. 2008; Lieberman et al. 2011).

A study has indicated that diverse severe human infections might be induced by different subtypes of *Mycobacterium tuberculosis* (MTB) as few symptoms are similar to other pathogens, e.g., nontuberculous mycobacteria (NTM) (Liu et al. 2014). A novel DNA barcoding visualization method has been exploited by this study to detect molecular typing of 17 mycobacteria genomes referred from the

database of NCBI prokaryotic genome database. The effective inter-strain pathogenic variations are represented from the detection of three mycobacterium genes (Rv3508, Rv3514, and Rv0279c) obtained from the MT Band's PE/PPE family (Liu et al. 2014). Similarly, many strains that were resistant after laboratory selection may provide in-depth knowledge of evolutionary mechanism for their new resistance gene adaptation from the environment or from other microbes (such as *Bacillus anthracis*) (Serizawa et al. 2010). Many bacteria codes are well-characterized toxins, which are known to cause severe diseases, e.g., HUS by EHEC (Kaper et al. 2004), streptococcus causing toxic shock syndrome (Lappin and Ferguson 2009), and *Corynebacterium diphtheriae* causing diphtheria (Holmes 2000).

5 Nanomaterial-Based DNA Barcoding

These innovative techniques identify a number of infectious pathogens quickly, which are present in trace amount. These pathogens with high sensitivity and specificity are in invariable demand to prevent the global morbidity and mortality. The spread of disease needs to be controlled for reducing the economic burden. Advancement in developing the fluorescent nanoparticle, super-paramagnetic nanoparticle, and metallic nanoparticle is very effective for bio-imaging detection of an infectious microorganism. This procedure is helpful in capturing the bacteria and viruses in solutions or biological samples (in vivo and in vitro) (Tallury et al. 2010). Fluorescent barcoded DNA assay is based on two nanoparticles (NPs): magnetic nanoparticles (MNPs) and gold nanoparticles (Au-NP) (Zhang et al. 2009). This helps in the rapid detection of the *Salmonella enteritidis* gene.

Recently, *Bacillus globigii* (Bg) spores simulate *B. anthracis* and other bacterial species. DNA barcoded with nanowires gold/silver/nickel (Au/Ag/Ni) is used for multiplexed detection of three antigens (Tok et al. 2006). However, large-scale synthesis of high-quality barcode nanowires is challenging. With the advancement of molecular imaging and multiplexed bioassays, drug discovery, gene profiling, and clinical diagnostics are made easy and cost-effective. This ultimately reduces the global burden of mortality rate and endemic spread (Lee et al. 2010). Intracellular bacteria are biodegradable and biocompatible in nature as they act as a good nonviral transfer of plasmid DNA into mammalian cells and surfaces to study the antigen-antibody recognition (Akin et al. 2007).

Another new technique called multicolor conjugated polymer nanoparticle (CPN) was developed by Feng et al. (2012). The bacteria and CPNs were self-assembled using a simple and time-saving manner. By tuning fluorescence resonance energy transfer (FRET) efficiencies, the CPN has shown production of assorted colors under single excitation wavelength among CPNs. Different excitation sources of flow cytometry and fluorescence microscopy have been matched by the wavelength showed up to 170 nm large Stokes shifts. It has also been applied for optical barcoding and cell imaging successfully. According to Feng et al. (2012), various barcodes with color-coded microparticles were prepared by mixing *E. coli* and the

CPNs together under one excitation. Four colors were used to prepare four different CPNs: green, yellow, blue, and red. It is observed that these colors show low toxicity toward cell and were implemented for cell imaging and optical barcoding, instigating new notion for presenting multifunctional structures and to self-assemble living organisms. These living organisms are present with functional materials and have a vast application in clinical research of water-soluble conjugated polymers in biomedical fields (Feng et al. 2012).

Effective control of *Mycobacterium tuberculosis* (MTB) is required to determine the subtype of this organism because their severe human infection symptoms will be induced by other MTB, for example, nontuberculous mycobacteria (NTM). This was the fastest method recommended on the basis of 16S rRNA gene and combination of three biomarker genes applied to other pathogens. Mycobacterium's three genes were presented by the entire genetic information, comprising a differentiating ability associated with 16S rRNA gene. An accurate molecular typing is obtained by 16S rRNA gene to bring trust for overcoming the disease in the future cases brought by *Mycobacterium* and to corroborate the genetic potentials (Liu et al. 2014).

6 Magnetic Barcoding

The detection of pathogenic microorganism on nucleic acid-based magnetic barcoding was introduced by Liong et al. (2013). A magnet was labeled over nanoprobe on microsphere that was specifically sequenced by using nuclear magnetic resonance technique. All the components were mixed up into single and small fluidic cartridges that were first streamlined to be used as a chip. This technology was used to detect drug resistance strains of *Mycobacterium tuberculosis* in sputum sample within 2.5 h. The versatility of cost-effective, advanced, and high-throughput technology along with magnetic barcode can be implemented to other diseases and researches. This technology could effectively indicate the biodefense and nosocomial infection by modifying the probe sequence (Liong et al. 2013). The effective genetic mutations in chronic diseases, including diabetes, heart diseases, and cancer, are studied by a bedside tool. It may also be used to shed some light for detecting epidermal growth factor receptor causing lung cancer by a single-point mutation in exon 21 (Dufort et al. 2011). The accessibility of particular technology is not restricted to magnetic readout, but it can be broadly used to plasmonic and luminescent studies based on gold nanoprobes and quantum dots, respectively (Hill and Mirkin 2006).

Wang et al. (2012) studied the pathogenicity pool of *Helicobacter pylori* by using this novel integrated technique. To apparently differentiate origin-specific genomic regions, the imaging method transforms frequency matrices to gray-scale levels to investigate *H. pylori* barcoded genome. In the National Centre for Biotechnological Information (NCBI) on prokaryotic genome database, six strains of H. pylori strains were compared with barcodes of *E. coli* O157: H7 strain. Wang et al. (2012)

submitted the following criteria to be adopted for recognizing the pathogenic potential of *H. pylori* (PAIs):

- Length is greater than 10,000 continuous base pairs.
- Barcode distance is distinct from that of the general background.
- Containing genes with known virulence-related functions (as determined by PfamScan and Blast2GO).

Very large DNA fragments (obtained through horizontal transfer) bearing multiple genes that encode virulence factors of bacteria are called PAIs (Schneider et al. 2011b). The heterogeneity is associated with great quality of biological systems. The main perceptive of the system is to develop a method, which uniquely identifies individual components and has interactions with each other. Peikon et al. (2014) developed a new method that tagged individual cell in vivo with genetic "barcode," which can be recovered by DNA sequencing. The viability of this method has been within the bacterial cells. Peikon et al. (2014) were the first to report an in vivo barcoding scheme with the perspective to label all of the individual cells of an entire tissue or organism. A general DNA barcode data processing system called "bio-barcode" was designed by Lim et al. (2009). This system encouraged quick gain in species DNA sequence database to meet global standards by providing specialized services. This makes barcoding inexpensive for Asian researchers.

7 Conclusion

The data of bacterial DNA barcoding is very limited and is not stored in one database to be easily accessible by the scientific community worldwide. Furthermore, the global initiative is required to facilitate the greeted potential of DNA barcoding tool to direct day-to-day infection control in hospitals and the community. The usefulness of DNA barcoding should be acknowledged, which is not only efficient in an identification of unnamed sequences, but it may also play an important role in the surveillance and control of microbial endemic diseases. For the designing of unspecified or unidentified bacterial diseases, treatment will be easier in the near future, if the researches are united and shared in a common platform. It can be done by developing a database; and all the researchers would deposit their sequence data for future reference. Therefore, this area of research is needed globally to counter the hazardous bacterial infections for future developments to control their spread.

The data available on bacterial DNA barcoding is limited and used for global bio-identification despite of all the criticism on DNA barcoding by the classical researchers (Rubinoff et al. 2006; Will et al. 2005). If a global system is developed for sharing the findings and interpreting the sequenced information, DNA barcoding becomes accessible in laboratory scale for maximizing benefit to the community. The comparison of assorted strains and the implementation of knowledge are based on the sharing and interpretation of data. In the first stage of global collaboration, all the information related to pathogens will be stored in details such as resistance

profiles and disease details especially for those that will be used as agents of bioterrorism. This is a great challenge for the scientist to form a global database of all the bacteria's sequences, which constitute hundreds and thousands of sequenced genomes annually. Moreover, the associated data, called metadata, will be available to the whole scientific community. It has been concluded that the comparison and sharing of results extracted in different laboratories can be facilitated by the harmonization of procedures such as data sequencing and data analysis.

Acknowledgment The authors are very thankful to all the associated personnel in any reference that contributed in/for the purpose of this research.

Conflict of Interest The authors declare no conflict of interest.

References

Akin D, Sturgis J, Ragheb K, Sherman D, Burkholder K, Robinson JP, Bhunia AK, Mohammed S, Bashir R (2007) Bacteria-mediated delivery of nanoparticles and cargo into cells. Nat Nanotechnol 2(7):441–449. https://doi.org/10.1038/nnano.2007.149

Auburn S, Campino S, Clark TG, Djimde AA, Zongo I, Pinches R, Manske M, Mangano V, Alcock D, Anastasi E, Maslen G (2011) An effective method to purify Plasmodium falciparum DNA directly from clinical blood samples for whole genome high-throughput sequencing. PLoS One 6(7):e22213. https://doi.org/10.1371/journal.pone.0022213

Balasingham SV, Davidsen T, Szpinda I, Frye SA, Tonjum T (2009) Molecular diagnostics in tuberculosis. Mol Diagn Ther 13(3):137–151. https://doi.org/10.1007/bf03256322

Ball SL, Armstrong KF (2006) DNA barcodes for insect pest identification: a test case with tussock moths (Lepidoptera: Lymantriidae). Can J For Res 36(2):337–350. https://doi.org/10.1139/x05-276

Barberan A, Casamayor EO (2010) Global phylogenetic community structure and β-diversity patterns in surface bacterioplankton metacommunities. Aquat Microb Ecol 59(1):1–10. https://doi.org/10.3354/ame01389

Begerow D, Nilsson H, Unterseher M, Maier W (2010) Current state and perspectives of fungal DNA barcoding and rapid identification procedures. Appl Microbiol Biotechnol 87(1):99–108. https://doi.org/10.1007/s00253-010-2585-4

Beheregaray LB (2008) Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. Mol Ecol 17(17):3754–3774. https://doi.org/10.1111/j.1365-294x.2008.03857.x

Bizzini A, Jaton K, Romo D, Bille J, Prod'hom G, Greub G (2011) Matrix-assisted laser desorption ionization–time of flight mass spectrometry as an alternative to 16S rRNA gene sequencing for identification of difficult-to-identify bacterial strains. J Clin Microbiol 49(2):693–696. https://doi.org/10.1128/jcm.01463-10

Blaxter ML (2004) The promise of a DNA taxonomy. Philos Trans R Soc Lond B Biol Sci 359 (1444):669–679. https://doi.org/10.1098/rstb.2003.1447

Croxatto A, Prod'hom G, Greub G (2012) Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. FEMS Microbiol Rev 36(2):380–407. https://doi.org/10.1111/j.1574-6976.2011.00298.x

Curtis TP, Sloan WT, Scannell JW (2002) Estimating prokaryotic diversity and its limits. Proc Natl Acad Sci USA 99(16):10494–10499. https://doi.org/10.1073/pnas.142680199

Didelot X, Bowden R, Wilson DJ, Peto TE, Crook DW (2012) Transforming clinical microbiology with bacterial genome sequencing. Nat Rev Genet 13(9):601–612. https://doi.org/10.1038/nrg3226

Dufort S, Richard MJ, Lantuejoul S, De Fraipont F (2011) Pyrosequencing, a method approved to detect the two major EGFR mutations for anti EGFR therapy in NSCLC. J Exp Clin Cancer Res 30(1):57. https://doi.org/10.1186/1756-9966-30-57

Feng X, Yang G, Liu L, Lv F, Yang Q, Wang S, Zhu D (2012) A convenient preparation of multi-spectral microparticles by bacteria-mediated assemblies of conjugated polymer nanoparticles for cell imaging and barcoding. Adv Mater 24(5):637–641. https://doi.org/10.1002/adma.201102026

Fenollar F, Puechal X, Raoult D (2007) Whipple's disease. N Engl J Med 356(1):55–66. https://doi.org/10.1056/nejmra062477

Greub G, Kebbi-Beghdadi C, Bertelli C, Collyn F, Riederer BM, Yersin C, Croxatto A, Raoult D (2009) High throughput sequencing and proteomics to identify immunogenic proteins of a new pathogen: the dirty genome approach. PLoS One 4(12):e8423. https://doi.org/10.1371/journal.pone.0008423

Hajibabaei M, Singer GA, Hebert PD, Hickey DA (2007) DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends Genet 23(4):167–172. https://doi.org/10.1016/j.tig.2007.02.001

Hebert PD, Cywinska A, Ball SL (2003) Biological identifications through DNA barcodes. Proc R Soc Lond B Biol Sci 270(1512):313–321. https://doi.org/10.1098/rspb.2002.2218

Hickerson MJ, Meyer CP, Moritz C (2006) DNA barcoding will often fail to discover new animal species over broad parameter space. Syst Biol 55(5):729–739. https://doi.org/10.1080/10635150600969898

Hill HD, Mirkin CA (2006) The bio-barcode assay for the detection of protein and nucleic acid targets using DTT-induced ligand exchange. Nat Protoc 1(1):324. https://doi.org/10.1038/nprot.2006.51

Hillemann D, Weizenegger M, Kubica T, Richter E, Niemann S (2005) Use of the genotype MTBDR assay for rapid detection of rifampin and isoniazid resistance in Mycobacterium tuberculosis complex isolates. J Clin Microbiol 43(8):3699–3703. https://doi.org/10.1128/jcm.43.8.3699-3703.2005

Holmes RK (2000) Biology and molecular epidemiology of diphtheria toxin and the tox gene. J Infect Dis 181(Suppl 1):S156–S167. https://doi.org/10.1086/315554

Hugenholtz P, Goebel BM, Pace NR (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol 180(18):4765–4774. https://doi.org/10.1016/0167-7799(96)10025-1

Iacono M, Villa L, Fortini D, Bordoni R, Imperi F, Bonnal RJ, Sicheritz-Ponten T, De Bellis G, Visca P, Cassone A, Carattoli A (2008) Whole-genome pyrosequencing of an epidemic multidrug-resistant *Acinetobacter baumannii* strain belonging to the European clone II group. Antimicrob Agents Chemother 52(7):2616–2625. https://doi.org/10.1128/aac.01643-07

Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, Wimalarathna H, Harrison OB, Sheppard SK, Cody AJ, Maiden MC (2012) Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. Microbiology 158(4):1005–1015. https://doi.org/10.1099/mic.0.055459-0

Kaper JB, Nataro JP, Mobley HL (2004) Pathogenic *Escherichia coli*. Nat Rev Microbiol 2(2):123–140. https://doi.org/10.1038/nrmicro818

Kuroda M, Sekizuka T, Shinya F, Takeuchi F, Kanno T, Sata T, Asano S (2012) Detection of a possible bioterrorism agent, *Francisella sp.*, in a clinical specimen by use of next-generation direct DNA sequencing. J Clin Microbiol 50(5):1810–1812. https://doi.org/10.1128/jcm.06715-11

Lappin E, Ferguson AJ (2009) Gram-positive toxic shock syndromes. Lancet Infect Dis 9(5):281–290. https://doi.org/10.1016/s1473-3099(09)70066-0

Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lund O (2012) Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol 50(4):1355–1361

Le T, Trexel J, Brugler A, Tiebel K, Maezato Y, Robertson B, Blum P (2013) Bar-coded Enterobacteria: an undergraduate microbial ecology laboratory module. Am J Educ Res 1(1):26–30. https://doi.org/10.12691/education-1-1-6

Lee H, Kim J, Kim H, Kim J, Kwon S (2010) Colour-barcoded magnetic microparticles for multiplexed bioassays. Nat Mater 9(9):745–749. https://doi.org/10.1038/nmat2815

Lieberman TD, Michel JB, Aingaran M, Potter-Bynoe G, Roux D, Davis MR Jr, Skurnik D, Leiby N, LiPuma JJ, Goldberg JB, McAdam AJ (2011) Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. Nat Genet 43(12):1275–1280. https://doi.org/10.1038/ng.997

Lim J, Kim SY, Kim S, Eo HS, Kim CB, Paek WK, Kim W, Bhak J (2009) BioBarcode: a general DNA barcoding database and server platform for Asian biodiversity resources. BMC Genomics 10(3):S8. https://doi.org/10.1186/1471-2164-10-s3-s8

Links MG, Dumonceaux TJ, Hemmingsen SM, Hill JE (2012) The chaperonin-60 universal target is a barcode for bacteria that enables de novo assembly of metagenomic sequence data. PLoS One 7(11):e49755. https://doi.org/10.1371/journal.pone.0049755

Liong M, Hoang AN, Chung J, Gural N, Ford CB, Min C, Shah RR, Ahmad R, Fernandez-Suarez M, Fortune SM, Toner M (2013) Magnetic barcode assay for genetic detection of pathogens. Nat Commun 4:1752. https://doi.org/10.1038/ncomms2745

Liu B, Zhang X, Huang H, Zhang Y, Zhou F, Wang G (2014) A novel molecular typing method of Mycobacteria based on DNA barcoding visualization. J Clin Bioinforma 4(1):4. https://doi.org/10.1186/2043-9113-4-4

Makarova O, Contaldo N, Paltrinieri S, Kawube G, Bertaccini A, Nicolaisen M (2012) DNA barcoding for identification of 'Candidatus Phytoplasmas' using a fragment of the elongation factor Tu gene. PLoS One 7(12):e52092. https://doi.org/10.1371/journal.pone.0052092

Mecsas J, Bilis I, Falkow S (2001) Identification of attenuated Yersinia pseudotuberculosis strains and characterization of an orogastric infection in BALB/c mice on day 5 postinfection by signature-tagged mutagenesis. Infect Immun 69(5):2779–2787. https://doi.org/10.1128/iai.67.5.2779-2787.2001

Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, Prior K, Szczepanowski R, Ji Y, Zhang W, McLaughlin SF (2011) Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104: H4 outbreak by rapid next generation sequencing technology. PLoS One 6(7):e22751. https://doi.org/10.1371/journal.pone.0022751

Miller SE (2005) Proposed standards for BARCODE records in INSDC (BRIs). Request document for continuation of support by the Alfred P. Sloan Foundation submitted by the Smithsonian Institution on behalf of Consortium for the barcode of Life: 22 January 2006, pp 36–38

Neigel J, Domingo A, Stake J (2007) DNA barcoding as a tool for coral reef conservation. Coral Reefs 26(3):487. https://doi.org/10.1007/s00338-007-0248-4

Peikon ID, Gizatullina DI, Zador AM (2014) In vivo generation of DNA sequence diversity for cellular barcoding. Nucleic Acids Res 42(16):e127–e127. https://doi.org/10.1093/nar/gku604

Quinones-Mateu ME, Avila S, Reyes-Teran G, Martinez MA (2014) Deep sequencing: becoming a critical tool in clinical virology. J Clin Virol 61(1):9–19. https://doi.org/10.1016/j.jcv.2014.06.013

Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, Paxinos EE, Sebra R, Chin CS, Iliopoulos D, Klammer A (2011) Origins of the E. coli strain causing an outbreak of hemolytic–uremic syndrome in Germany. N Engl J Med 365(8):709–717. https://doi.org/10.1056/nejmoa1106920

Ratnasingham S, Hebert PD (2007) BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). Mol Ecol Resour 7(3):355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Rhode H, Qin J, Cui Y, Li D, Loman NJ, Hentschke M, Chen W, Pu F, Peng Y, Li J, Xi F, E. coli O104: H4 Genome Analysis Crowd-Sourcing Consortium (2012) Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104: H4. N Engl J Med 365(8):718–724

Ripa T, Nilsson PA (2007) A Chlamydia trachomatis strain with a 377-bp deletion in the cryptic plasmid causing false-negative nucleic acid amplification tests. Sex Transm Dis 34(5):255–256. https://doi.org/10.2807/esw.11.45.03076-en

Rubinoff D, Cameron S, Will K (2006) A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. J Hered 97(6):581–594. https://doi.org/10.1093/jhered/esl036

Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopaedia of life: an introduction to DNA barcoding. Philos Trans R Soc Lond B Biol Sci 360 (1462):1805–1811. https://doi.org/10.1098/rstb.2005.1730

Schneider G, Dobrindt U, Middendorf B, Hochhut B, Szijarto V, Emődy L, Hacker J (2011a) Mobilisation and remobilisation of a large archetypal pathogenicity island of uropathogenic *Escherichia coli* in vitro support the role of conjugation for horizontal transfer of genomic islands. BMC Microbiol 11(1):210. https://doi.org/10.1186/1471-2180-11-210

Schneider KL, Marrero G, Alvarez AM, Presting GG (2011b) Classification of plant associated bacteria using RIF, a computationally derived DNA marker. PLoS One 6(4):e18496. https://doi.org/10.1371/journal.pone.0018496

Serizawa M, Sekizuka T, Okutani A, Banno S, Sata T, Inoue S, Kuroda M (2010) Genomewide screening for novel genetic variations associated with ciprofloxacin resistance in Bacillus anthracis. Antimicrob Agents Chemother 54(7):2787–2792. https://doi.org/10.1128/aac.01405-09

Tallury P, Malhotra A, Byrne LM, Santra S (2010) Nanobioimaging and sensing of infectious diseases. Adv Drug Deliv Rev 62(4):424–437. https://doi.org/10.1016/j.addr.2009.11.014

Tanabe AS, Toju H (2013) Two new computational methods for universal DNA barcoding: a benchmark using barcode sequences of bacteria, archaea, animals, fungi, and land plants. PLoS One 8(10):e76910. https://doi.org/10.1371/journal.pone.0076910

Tok JBH, Chuang F, Kao MC, Rose KA, Pannu SS, Sha MY, Chakarova G, Penn SG, Dougherty GM (2006) Metallic striped nanowires as multiplexed immunoassay platforms for pathogen detection. Angew Chem Int Ed 45(41):6900–6904. https://doi.org/10.1002/anie.200601104

Wang J, Soininen J, He J, Shen J (2012) Phylogenetic clustering increases with elevation for microbes. Environ Microbiol Rep 4(2):217–226. https://doi.org/10.1111/j.1758-2229.2011.00324.x

Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. Syst Biol 54(5):844–851. https://doi.org/10.1080/10635150500354878

Wolk DM, Struelens MJ, Pancholi P, Davis T, Della-Latta P, Fuller D, Picton E, Dickenson R, Denis O, Johnson D, Chapin K (2009) Rapid detection of *Staphylococcus aureus* and methicillin-resistant *S. aureus* (MRSA) in wound specimens and blood cultures: multicenter preclinical evaluation of the Cepheid Xpert MRSA/SA skin and soft tissue and blood culture assays. J Clin Microbiol 47(3):823–826. https://doi.org/10.1128/jcm.01884-08

Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, Hooper SD, D'haeseleer P, Zemla A, Singer M, Lapidus A, Nolan M, Copeland A, Han C, Chen F, Cheng JF, Lucas S, Kerfeld C, Lang E, Gronow S, Chain P, Bruce D, Rubin EM, Kyrpides NC, Klenk HP, Eisen JA (2009) A phylogeny-driven genomic encyclopaedia of bacteria and archaea. Nature 462:1056–1060. https://doi.org/10.1038/nature08656

Zhang D, Carr DJ, Alocilja EC (2009) Fluorescent bio-barcode DNA assay for the detection of *Salmonella enterica* serovar Enteritidis. Biosens Bioelectron 24(5):1377–1381. https://doi.org/10.1016/j.bios.2008.07.081

Zubair M, Malik A, Ahmad J (2012) Study of plasmid-mediated extended-spectrum β-lactamase-producing strains of enterobacteriaceae, isolated from diabetic foot infections in a North Indian tertiary-care hospital. Diabetes Technol Ther 14(4):315–324. https://doi.org/10.1089/dia.2011.0197

Part III
DNA Barcoding in Plants

DNA Barcoding: Implications in Plant-Animal Interactions

Muniyandi Nagarajan, Vandana R. Prabhu, Ranganathan Kamalakkannan, and Palatty Allesh Sinu

Abstract Trophic interactions between plants and animals occur through various ecological processes such as pollination, frugivory, and herbivory. Among the three, pollination and frugivory represent the mutualistic interactions and the herbivory represents a negative antagonistic interaction. Interaction with pollinators and frugivores are statutory for many plants for the dispersal of gametes and offspring, respectively. On the contrary, the herbivory plays a substantial role in maintaining the stability of floral communities. Disruption in these natural species interactions can have detrimental effect on both the interacting species and other species of higher trophic levels, which eventually destabilizes the terrestrial ecosystem. Thus, a thorough perception of the interacting species is imperative for the restoration of bio-communities and conservation of ecosystems. Although, plant-animal interactions had been identified traditionally through field observation and tracking the movement of plant visitors, they are inadequate in accurately determining the interacting species as interaction occurs in the vicinity of other co-occurring plants and animals. Similarly, histological and biochemical analyses of gut matter and feces of herbivores/frugivores are not adequate to accurately identify the plant sources that the animals are depending upon. Paucity of an appropriate methodology has been always a limitation in delineating full range of plant-animal interactions. DNA barcoding technique has revolutionized the field of community ecology. Its unprecedented accuracy and speed has made it, ideally the best method of choice for identifying plant-animal interactions at species and population levels. DNA barcoding regions *rbcL*, *matK*, *trnH-psbA*, *trnL*, ITS2 (plant), and COI (animal) are the most common markers used in identifying plant and animal species. These barcodes have been adapted to characterize various plant and animal sources (honey,

M. Nagarajan (✉) · V. R. Prabhu · R. Kamalakkannan
Department of Genomic Science, School of Biological Sciences, Central University of Kerala, Kasargod, Kerala, India
e-mail: nagarajan@cukerala.ac.in

P. A. Sinu
Department of Animal Science, School of Biological Sciences, Central University of Kerala, Kasargod, Kerala, India

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_8

pollen, gut content, and feces) for unraveling many of the unnoticed trophic associations existing between plants and animals.

Keywords Plant-animal interaction · Pollen · Honey · Seed · Frugivore · DNA barcoding

1 Introduction

Interactions of species build up the food web in any given ecosystem. A stable community has a complex species interaction between and among plants, plants and animals, and animals and animals. The abiotic factors such as water, soil, temperature, energy, etc. facilitate the biotic interactions in a community. Plants and animals occupy a major share of ecological communities. A natural community is understood well through studying almost all possible biotic interactions above and below ground, which includes plant-microbial interactions and plant-animal interactions. While mutualism is a facilitative plant-animal interaction, herbivory is an antagonistic plant-animal interaction. However, these two interactions are vital to maintain stability in any given terrestrial community and ecosystem.

Pollination networks play a major role in community dynamics in which pollinators interact positively with floral populations. The relationship mainly involves offering food (pollen or nectar) by plants to the pollinators as a reward paid for their service in mediating pollination. Such an important service provided by pollinators serves as the most decisive facet of reproduction in flowering plants and also contributes remarkably to floral evolution and plant adaptations, eventually leading to speciation. Similarly, frugivore-plant relationship remains another classic example of mutualistic interactions. Frugivores (fruit eaters) are usually involved in a process of spatial distribution of the seeds and thus are vital for the persistence of floral populations. Besides the mutualistic interactions, antagonistic interactions also occupy an important place in maintaining stable ecological communities. A major portion of the plant biomass is good source of food for a tremendous array of herbivores, comprising the second trophic level of food chain. Such negative interactions existing between plants and herbivores are considered as one of the chronic causes for plant damage. Though herbivores have a deleterious effect on the plant communities, in an ecological viewpoint, they act as major regulators of ecosystem functions.

Regardless of the ways in which a species interact with others, most trophic level interactions seen between plants and animals are species specific. For instance, in the pollination networks, a few pollinators manifest foraging preferences toward specific flowering plants. This is substantiative in the case of orchids, where nearly 60% of the orchid families are pollinated by pollen-collecting wasps (Mchatton 2011). Similar cognate relationship has been noted between fruits and frugivores. A number of fruiting plants require intervention of specific animal vectors for the dispersal of seed and its successful germination. The loss of Calvaria forest due to the extinction

of dodo bird is a classic example of such species-specific ecological relationship. Likewise, in the case of antagonistic interactions, specifically in the plant-herbivore interactions, most herbivores have a narrow diet range. Though a few plant-animal interactions have been pointed out, the accurate feeding behaviors and foraging preferences of many herbivores, frugivores, and pollinators are still obscure. The organismal interactions between plants and primary consumers, being the fundamental unit on which other higher trophic level interactions are built, need to be thoroughly studied.

Notwithstanding the ecological importance of plant-animal interactions, their identification has always been a challenge. Several attempts have been made to study the animal and plant species involved in various trophic interactions. Earlier, identification of ecological relationships was based on traditional methods like field observation, microscopic analysis, tracking animals using traps and dyes, etc. Other indirect approaches have included various histological and biochemical procedures that target on analyzing the gut or fecal matter to identify the plant materials. Although researchers detect trophic level connections between plants and animals through conventional methods, this is not a choice under many situations, such as when they interact in the vicinity of many co-occurring plant and animal species or when the ingested diet meals of primary consumers cannot be easily discerned. Such circumstances require a suitable alternative improved and accurate method of identification processes. In the recent years, DNA-based method has received much attention among molecular taxonomists as it allows identification of organisms up to species and the population levels. Specifically, DNA barcoding has proven to be a good candidate for species identification as it uses variable regions in the genome that clearly differs even between closely related species. This chapter focuses on various kinds of plant-animal interactions and the usefulness of DNA barcoding in understanding the trophic relationships between plants and animals.

2 Plant-Animal Interplay: An Ecological Outlook

The first evidence of plant-animal coevolution was reported by Charles Darwin by relating the morphologic specificity between orchids and moths (Darwin 1859, 1862). The present concept of plant-animal evolutionary associations results from his views. Plant-animal interactions (both mutualistic and antagonistic associations) are the major driving force for plant-animal coevolution (Clare 2014). Exchange of genetic content between the interacting species is assumed to be zero in plant-animal evolutionary interactions (Ehrlich and Raven 1964; Clare 2014). It is through the food chain networks that most of the plant species forge relationship with pollinators, frugivores, and herbivores (Fig. 1). Therefore understanding the connections between plants and animals with an ecological perspective is very essential to accurately determine their status in the food chain and to conserve the endangered plant or animal species.

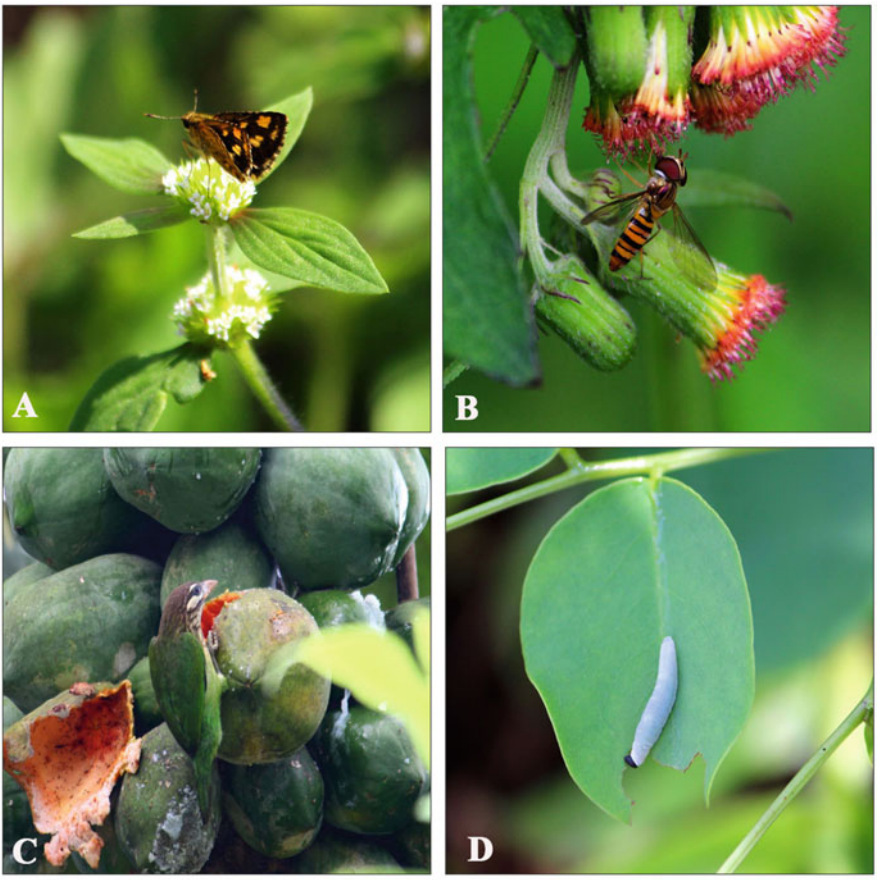


Fig. 1 Different forms of plant-animal interactions. (**a** and **b**) Plant and pollinator. (**c**) Plant and frugivore. (**d**) Plant and herbivore

2.1 Plants and Pollinators

Flowers are the most attractive part of angiosperms that draw attention of many insects and birds by their color, scent, and shape alone or in combination in order to carry out pollination and set seeds (Ollerton et al. 2011; Schiest and Johnson 2013). In return, the flowers provide the pollinators with food in the form of pollen and/or nectar for their priceless service. The process of pollen transfer is mainly mediated by wind (anemophily) and animals (bees, butterflies, birds, etc.). However, besides the anemophilic mode of pollination, a great diversity of flowering plants depends on animals for reproduction (Ollerton et al. 2011). The ecological importance of pollinators, therefore, is much higher in the natural communities as their mutual dependence on the flora impacts persistence of both plant and animal populations (Hegland et al. 2009).

Several studies have been emphasized on the evolution of plant-pollinator interaction, food habits of major pollinators, and their role in stabilizing natural communities and conserving ecosystem. Sargent and Ackerly (2007) described three different processes related to plant-pollinator interactions—habitat filtering, diffuse facilitation, and competitive exclusion—to explain which species establish and persist. A recent study by Robson (2014) postulated the importance of wild insect pollinators on crop yield. Wild insects normally prefer wild flowers as they provide sufficient quantity of nectar. The study ranked 41 wild species that share common pollen vectors with canola (*Brassica napus L.*) a cultivated crop (which does not provide optimal resource to their visitors), suggesting the usefulness of incorporating selected wild flowering plants in the crop land for attracting important pollinators and increasing crop production. However, it is important to notice that the foraging behavior of the pollinators, the species visited in a given visitation bout and the diversity of the pollen carrying on the body of the pollinators influence the pollination efficiency of the pollinators.

2.2 Plants and Frugivores

Fruit-frugivore interaction is one of the chief components of community ecology that plays a significant role in plant restoration and radiation (Chama et al. 2013). Frugivores (fruit eaters) are natural seed disseminators as they involve in a fundamental spatial process of dispersing propagules to long and short distances. Such spatial patterns of seed dissemination influence the rate of gene migration within and among plant populations and are subjected to post-dispersal processes like selection and predation, eventually affecting persistence of plant populations (Robledo-Arnuncio and Garcia 2007). Nearly 90% of the plant species bearing fleshy fruits heavily depend on animals to carry out the process of seed dissemination across space and time (Chama et al. 2013). However, it cannot be generalized that frugivores are always in mutualistic association with plants. The seed handling behavior of the animals determines the role played by the animals, whether to be a seed disperser or seed predator (Chama et al. 2013).

Fruit-frugivore interactions are widely studied to get a better understanding of frugivore diet habits, evolution of fruit traits, and plant regeneration dynamics. Encinas-Viso et al. (2014) using an individual-based model studied the diversification of fruit traits. The reason for such evolution of trait was predicted as seed dispersion by frugivores, their foraging behavior and numerical abundance. Another attempt was made by Sankamethawee et al. (2011) to find out the hinge between plants bearing small fleshy fruits and avian frugivores in an evergreen forest of northeast Thailand. Over 53 species of birds, most commonly fairy-bluebirds, barbets, and bulbuls, were reported to be interacting with nearly 136 plant species. Majority of these interactions were not specific, except Singkrang (*Saurauia roxburghii*) and figs, which showed tight link with flowerpeckers and thick-billed pigeons, respectively, for seed dispersal.

2.3 Plants and Herbivores

Plant-herbivore interactions are one of the most recognized plant-enemy interactions, which play a central role in regulating the functions of ecosystems (Schallhart et al. 2012). Herbivores are primary consumers that consume plant biomass (a process known as herbivory) and cause damage to plant populations by reducing plant survival and fitness. Effects of such negative interaction in limiting the size of plant population are higher than intra-/interspecific competitions existing among plant species (Wirth et al. 2008). Most of the members of class Insecta are herbivores that are adapted to eat plants (Jurado-Rivera et al. 2009). These phytophagous insects illustrate one of the major channels of energy that pass to various trophic levels through food chain (Garcia-Robledo et al. 2013). Herbivores are classified into generalists and specialists based on what they consume. A generalist feeds on a wide variety of plants, whereas diet of a specialist includes selected plant species. The ungulate vertebrates are the generalist herbivores, while the phytophagous insects are mostly specialist herbivores.

Investigation of herbivores' diet habits is more challenging since their relationship with plants are temporally and spatially dynamic. Feeding choices of herbivores depend on diversity and identity of plants. Schallhart et al. (2012) made an attempt to study the impact of floral identity as well as diversity on food preference of subterranean larvae *Agriotes* by providing different combinations and diversity of food plants such as maize, legume, grass, and forb to them. The results indicated that plant diversity plays a major role in regulating dietary choices of herbivores as the larvae changed their feeding preferences in accordance with increased plant diversity. Though plants are exposed to herbivory, a few plants develop defense mechanisms against this negative interaction. The widespread ant-plant mutualistic association serves as a good example of such defense mechanism. Ants protect their host plants by predation of herbivores and their eggs, and, in return, plants provide housing and extrafloral nectar for the ants to ensure their perpetual protection against herbivory (Scharmann et al. 2013).

3 DNA Barcoding in Plant-Animal Interactions

Pollinators, frugivores, and herbivores manifest foraging preferences for multiple plant species. Therefore, understanding the food habits of these animals is critical with respect to trophic ecology as they reveal most of the difficult to observe trophic connections between plants and animals in the natural communities. DNA barcoding technique has opened up opportunities to identify plant species from various samples such as pollen, honey, digested plant parts, feces, etc. with utmost precision (Wilson et al. 2010; Nakashima et al. 2010; Gonzalez-Varo et al. 2014; Bruni et al. 2015; Nagarajan et al. 2016; Galimberti et al. 2016). The mitochondrial gene COI, which is extensively being used to differentiate animal species, evolves far too slowly in

plants (Hebert et al. 2003; Cowan and Fay 2012; Fazekas et al. 2012; Li et al. 2014; Ali et al. 2015). Because of such lower evolutionary rates, mtDNA markers hold little potential in discriminating plant species (Li et al. 2014). As alternatives, various coding and noncoding regions of plastid genome have been used as an effective target for barcoding plant species. However, none of these loci work across all the species due to introgressive hybridization and polyploidy events (Li et al. 2014; Ali et al. 2015).

The consortium for the barcode of life-plant working group (CBOL) has suggested a few plastid regions such as *rbcL*, *matK*, *trnH-psbA*, *trnL*, and the nuclear ribosomal internal transcribed spacer (ITS2) for the identification of plant species (Bruni et al. 2015). Though *matK* sequence shows higher mutation rate, it is less feasible in analyzing complex genome since it requires different primer combinations to amplify various angiosperm families (Hilu and Liang 1997; Dunning and Savolainen 2010). The other noncoding regions such as ITS2 and *trnH-psbA* have been reported as good candidates for discriminating congeneric plant taxa (Hollingsworth et al. 2011). These plant DNA barcoding regions remain as the most successful tool for studying plant-pollinator, plant-frugivore, and plant-herbivore interactions (Boyer et al. 2016). With the aid of such markers, a growing body of research has been conducted to identify plant species using DNA barcoding technique from honey, pollen, and feces for unraveling many of the unnoticed trophic associations existing between plants and animals.

3.1 Identifying the Provenance of Pollen and Honey

Pollen and honey are important study subjects for ecologists as it is possible to deduce their botanical constituents. Identification of source constituents of honey and pollen collected from distinct pollinators is fundamental to understand the process of pollen transfer and pollinator-plant interactions. Although traditional methods such as microscopy, direct field observation, and tracking of pollen movements by dyes and traps have been implemented, they lack taxonomic precision in determining the provenance of pollen and honey (Waser and Price 1982; Greenwood 1986; Linhart et al. 1987; Caron and Leblanc 1992; Parra et al. 1993; Richardson et al. 2015). In the recent years, literatures that have adopted DNA barcoding method (Fig. 2) to identify the floral source of pollinators from the honey and the pollen have grown up (Wilson et al. 2010; Schnell et al. 2010; Jain et al. 2013; Bruni et al. 2015; Galimberti et al. 2014; Wong et al. 2015).

In most of the plant-pollinator interactions, pollinators manifest foraging preferences for a vast variety of flowers. However, despite the wide-ranging plant-pollinator interactions, relatively, a few pollinators forge obligatory relationship with specific plant species (Kearns et al. 1998). The specificity of pollinator is conspicuous in the case of *Hylaeus* bees that are endemic to Hawaiian Island. Wilson et al. (2010) delineated that *Hylaeus* bees pollinate only a few native plant species such as *Dubautia menziesii*, *Leptecophylla tameiameiae*, and *Argyroxiphium sandwicense*

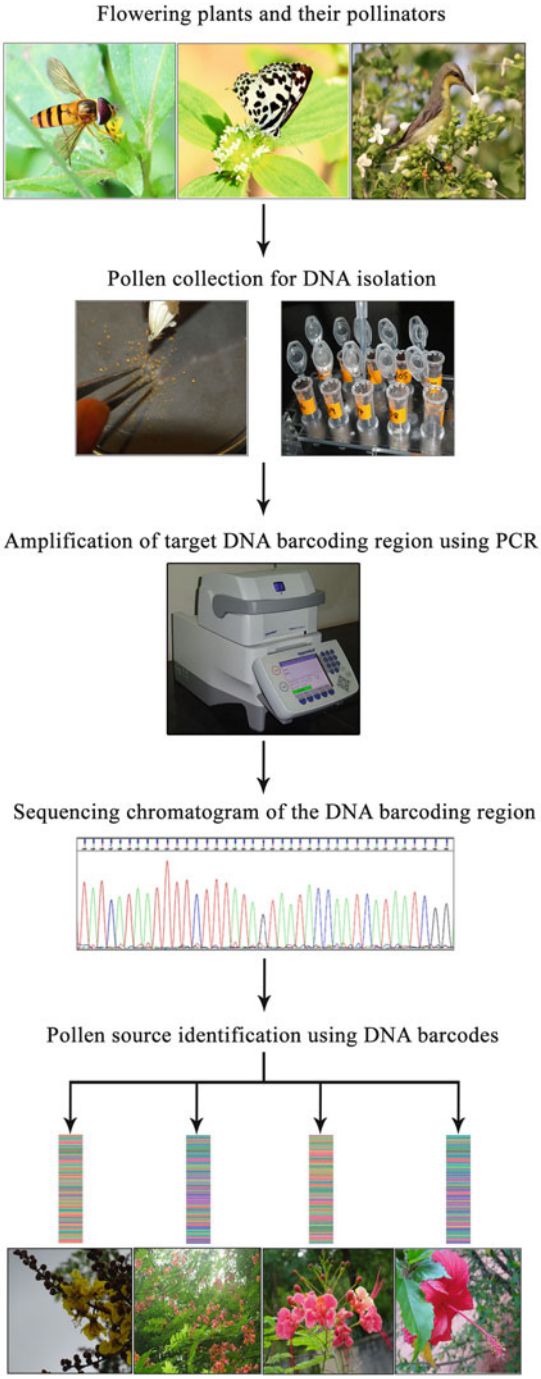

Flowering plants and their pollinators
Pollen collection for DNA isolation
Amplification of target DNA barcoding region using PCR
Sequencing chromatogram of the DNA barcoding region
Pollen source identification using DNA barcodes

Fig. 2 Schematic view of DNA barcoding approach to study the plant-pollinator interactions

in the vicinity of other co-occurring plants at flowering stage that include both native and a few alien species. The floral choices of Hylaeus bees inhabiting in different geographical locations of the island were identified by characterizing ITS and 5.8S *rDNA* regions of the pollen samples. All pollen grains collected from the bees belonged only to native plants, and among the native pollen, Hylaeus bees carried significantly large amount of pollen of an endangered species *Argyroxiphium sandwicense*, the commonly known Hawaiian silversword. It indicated that *Hylaeus* bees are important in mediating pollination of native plants and are the key pollinators of endangered silversword species. Recently, a similar study was conducted by Galimberti et al. (2014) using *rbcL* and *trnH-psbA* regions in order to identify the source plants of pollen collected from beehives, which resulted in the identification of 52 source plant species. Though both *rbcL* and *trnH-psbA* have been employed, *rbcL* region did not show much variation between closely related species in comparison with *trnH-psbA*.

Besides, characterizing pollen for unraveling the trophic connections between plants and pollinators, a few studies have identified potential pollinators using mitochondrial markers. For example, Wong et al. (2015) have characterized the pollinators associated with *Persicaria chinensis* using mitochondrial COI gene marker. *Persicaria chinensis*, the commonly known Chinese knotweed, are high-risk invasive species that are of great ecological and economic importance. Despite the economic and ecological value, it serves as a prime example of heterostyly, a structural modification in flowers that turn down self-pollination (Reddy et al. 1977). Such structural modifications of flowers along with strong scent and nectar imply high chance of cross pollination by insect vectors in these plants. However, previous studies have reported only a few insect vectors that facilitate pollination in this invasive species (Nishihiro and Washitani 1998; Thomas et al. 2009). Molecular characterization of the floral visitors of Chinese knotweed identified 4 orders and 23 distinct insect species as pollinators of this high-risk invasive species. In general, lack of pollinators does not act as a limiting factor for the propagation of invasive species (Richardson et al. 2000; Bartomeus et al. 2008). The detection of more than 20 distinct pollinator species for Chinese knotweed supported the previous studies. Regardless of the invasive nature of these species, the usefulness of incorporating this invasive species in the cultivation lands would promote pollination of important native plants.

The interaction among nectarivorous insects, specifically honeybees and plants, is easily discerned by characterizing the pollen contained in the honey. Although honeybees store pollen quite separately in their hives, a few infiltrate into the nectar and eventually get deposited in the honey. In many cases, the provenance of honey has been determined from the infiltrate using various genes such as *rbcL*, *trnH-psbA*, ITS, *adh1*, and COI (Schnell et al. 2010; Jain et al. 2013). For instance, Bruni et al. (2015) identified 39 plant species from honey samples.

3.2 Tracking the Seed Dispersal and Disperser

Because plants are sessile, they harness a wide range of animals for dispersing seeds. Therefore, their interaction with fruiting plants needs to be extensively studied to better understand their trophic relationships and seed dispersal events. At many instances, traditional methods face limitations and challenges in resolving fruit-frugivore interactions mainly due to thorough mastication and digestion of ingested plant parts such as seeds, pulp, or skin. DNA barcoding technique has alleviated these limitations and has rendered a new platform to precisely disclose plant-frugivore connections (Hayward 2013; Viana et al. 2015). In the recent years, use of DNA barcodes to investigate the feeding ecology of various frugivores has become very popular as they ultimately help to understand plant-frugivore relationships.

The potential of DNA barcodes has been examined by Galimberti et al. (2016) in identifying diet constituents of a frugivorous bird, *Sylvia atricapilla*, in a protected region of Northern Italy. By using the plastid DNA markers *rbcL* and *trnH-psbA*, the unscathed seeds and plant parts collected from the bird droppings were analyzed. Comparison of the barcode sequences with the reference DNA sequences of local plants resulted in the identification of 17 plant species in the diets of *Sylvia atricapilla*. Likewise, long distance dispersal (LDD) of seeds has also been identified with the support of plant barcodes. The first empirical report on such possibility of quantifying LDD was provided by Viana et al. (2015). The study analyzed the intact seeds collected from the gizzard of birds that were migrating from mainland to remote oceanic islands, with the aid of *matK* and *rbcL* markers and seed morphology. Characterization of these barcodes resulted in the identification of 21 different plant species providing evidence that migratory birds are capable of transporting seeds over a long distance during their flight. Further, most of the seeds transported did not appear in the remote islands, which suggests the significance of post-dispersal filters in impeding the establishment of plant species there. Similar to birds, volant species of mammals are also seen playing important role in the dispersal of seeds (Muscarella and Fleming 2007; Silveira et al. 2011). An investigation of fruit feeding habits of more distantly related Jamaican bats using *rbcL* markers resulted in the identification of 11 plant taxa from 135 fecal samples (Hayward 2013). But, only four taxa were identified when traditional method was implemented (Giannini and Kalko 2004; Lopez and Vaughan 2007; Teixeira et al. 2009).

On the other hand, a few studies have used mitochondrial DNA markers to reveal animal identities from the fecal or regurgitated samples (Nakashima et al. 2010; Gonzalez-Varo et al. 2014). Common palm civets are widely distributed small mammals of Southeast Asia. Their excessive dependence on fruits and ability to disperse larger seeds over a long distance mark them as important contributors of the ecological process (Davis 1962; Gruezo and Soligam 1990; Joshi et al. 1995; Su and Sale 2007). In general, large seeds appear to be a limiting factor for a number of dispersing agents (Corlett 1998). Civets are one among those few frugivores that disseminate large seeds over a long distance. Formerly, civet species were identified by examining the color, smell, size, shape, etc. of their fecal samples (Sinu et al.

2016). Since distinct civet species have similar body size and hair color, these identification strategies fail to accurately distinguish among several civet species. Nakashima et al. (2010) made an attempt to find out the identity of different civet species from the fecal samples using cytochrome *b* gene. The results indicated that most of the analyzed samples are of common palm civets. Also, seeds present in the feces were measured for determining their size and subsequently allowed to germinate to check its viability as well as to confirm the species. Seeds of fig plant, *Leea aculeata* and *Endospermum diadenum*, were most common in the civet feces, and among all the identified plant species, seeds of *Aglaia grandis* were the largest which provided a direct example of the contribution of common palm civets in large seed dispersal. Likewise, Gonzalez-Varo et al. (2014) made an attempt to find out bird dispersed seeds by characterizing 464 bp COI gene region. The study detected five frugivorous bird species from regurgitated/defecated seeds (belonging to four fleshy-fruited plant species) and postulated that the identified bird species are potentially involved in dispersing seeds of a few fleshy-fruited plants.

3.3 Identification of Herbivore Host Plants

Plant-herbivore network plays an essential role in regulating the functions of ecosystem (Schallhart et al. 2012). Insects alone comprise a major group of herbivores and consume large portion of plant biomass. When taken together, plants and the insects associated with them represent at least 50% of the total known species on earth (Futuyma and Agrawal 2009). Determining the plant-insect trophic link through traditional methods like microscopy, field observation, carbon isotope analysis, etc. is challenging as it is expensive and laborious and thus provides little reliable information about herbivore trophic dynamics. An increasing number of studies have made use of DNA barcodes, over the past few years, to investigate insect-host plant associations.

Matheson et al. (2008) tested the possibility of identifying the species of plant materials from insect's gut using 157 bp *rbcL* fragment. Members of eight distinct insect families were collected along with their host plants, and they were allowed only to follow a monophagous feeding mode to mitigate taxonomic ambiguity that would arise when multiple plant species are obtained from the guts. The amplification and subsequent sequencing of *rbcL* region identified the plant meal of distinct insects. The study also postulated that plant DNA can be even recovered from insect gut 12 h after ingestion.

Among the diverse families of insects, beetle family has been reported as one of the most abounding insect families, with much of its diversity having been emerged from plant-insect coevolution (Farrell 1998; Oberprieler et al. 2007). Diet habits of beetles and weevils have been widely studied. Jurado-Rivera et al. (2009) studied the host plant-chrysomelid beetle (leaf beetle) interaction by analyzing whole beetle extract using chloroplast *trnL (UAA)* intron region which resulted in the identification of 13 plant families that are consumed by about 76% species belonging to

Chrysomelinae family. The study also reported the preference of beetles for Myrtaceae and Fabaceae families. Likewise, molecular characterization of *trnL* region confirmed 26 plant families from the guts of weevils of tropical forest (Pinzón-Navarro et al. 2010). However, as an improvement over these studies, Kitson et al. (2013) tested the efficiency of *trnL (UAA)* barcode (by amplifying the region with different set of primers) in detecting plant species consumed by two closely related species of weevil, *C. murinus* and *C. ovalis.* In general, tropical herbivores (order, Coleoptera) feed on distantly related plant species, which is not concordant with phylogenetic conservatism (Novotny et al. 2002; Ødegaard et al. 2005). The results also revealed that weevil species follow a broad palatable diet, which includes a wide variety of plant species. Quite recently, feeding range of rolled-leaf beetles has been recorded by García-Robledo et al. (2013) with the aid of three loci (*rbcL*, *trnH-psbA*, and ITS2). Among the three loci, good quality sequences were generated only for *rbcL* and ITS2. The *rbcL* locus determined family of a few host plants, and, in some other cases, it identified the plants up to genus level, whereas ITS2 region was successful in the identification of host plants up to species level. A couple of years later, Kajtoch (2014) surveyed the ecological interaction between phytophagous beetle and their host plants through amplification of *trnL* and *rbcL* markers, followed by Sanger sequencing and Illumina sequencing for monophagous and polyphagous groups, respectively. The study revealed 224 host plant-beetle interactions, which provided inferences on the feeding habits of a few insect genera. Of these 224 interactions, a few beetles belonging to 7 distinct genera were found feeding on similar plant species. However, a much broad range of feeding interest was observed for the other four genera.

Molecular analysis of gut contents of soil-dwelling insects that feed on below-ground parts of plant has also been researched with the aid of plant-based markers such as *rbcL*, *trnL*, etc. (Staudacher et al. 2011; Wallinger et al. 2013). Feeding ecology of such subterranean insects would be useful in identifying the insect-plant association in the soil and major soil pests. The herbivory is not limited only with the feeding of plant parts but using the plant parts for other purposes, such as making a dwell and lining a brood cell, or for cultivating fungal bodies, which become the food of insects (Maclvor 2016; Kambli et al. 2017). Plant-leafcutter bee interaction is an exceptional case of this kind of herbivory. The conventional method to identify the preferred plants of the bees is the direct observation of bees cutting leaves from plants (Kambli et al. 2017). While this might be a tedious part of the field work, the DNA barcoding method can provide an accurate list of preferred plant species through sequencing the DNA extracted from the leaf discs, which are collected from the bee nests (Maclvor 2016).

4 Reconstructing Plant-Animal Link Through NGS

DNA barcoding, while being easy, less time-consuming, and able to distinguish organisms at species level, has low efficiency when applied to samples containing a mixture of plant species (Wilson et al. 2010; Hawkins et al. 2015). It hugely relies on databases to identify the plant species, by comparing the DNA sequences obtained in the studies with the reference sequences of databases. This also limits the efficiency of identification process because not all the plant barcode sequences have been deposited in the genetic databases. In addition to these limitations, traditional Sanger sequencing method, which is extensively being used to sequence the DNA barcodes, faces other serious pitfalls such as requirement of an additional cloning step for multiple amplicon sequencing and failure to generate sequences for all samples at once (Pompanon et al. 2012; Kajtoch 2014). Recent metabarcoding technology has succeeded in dealing with the problems faced by DNA barcoding method (Harismendy et al. 2009; Shokralla et al. 2012). This expeditious method integrates both the DNA-based identification and next-generation sequencing and allows the users to analyze complex matrices (i.e., honey) containing DNA of different plant species in a single run (McClenaghan et al. 2015).

Diet assessment of several herbivorous animals such as birds, insects, mollusks, ruminants, and other mammals has been carried out using metabarcoding (Soininen et al. 2009; Valentini et al. 2009; Pegard et al. 2009; Kowalczyk et al. 2011; Raye et al. 2011; Pompanon et al. 2012; Kajtoch 2014; Srivathsan et al. 2014; McClenaghan et al. 2015; Keller and Steffan-Dewenter 2014). Kajtoch (2014) provided the first and more accurate report on the feeding habit of polyphagous weevil, *Centricnemus leucogrammus*, by comparing the efficiency of both high-throughput sequencing and Sanger sequencing for *rbcL* and *trnL* regions. As expected, next-generation sequencing technique was found extremely suitable in analyzing dietary constituents of the beetle species and resulted in identifying 30 plant genera, whereas Sanger sequencing procedure was successful in delineating only 20 plant genera. Out of the two plant DNA region used, *rbcL* worked better in NGS. Sequencing success of *trnL intron* was also found slightly higher in NGS than the other. By contrast, in the case of Sanger sequencing method, *trnL* region identified more plant species. Similarly, McClenaghan et al. (2015) studied the dietary constituents of four grasshopper species inhabiting in the same area using *rbcL* marker through Illumina sequencing technology, which provided an explicit idea about the feeding behavior and resource partitioning among these sympatric grasshopper species. Furthermore, the dietary components of two leaf-feeding monkeys from fecal samples using P6 loop of *trnL* region have also been investigated (Srivathsan et al. 2014).

In the recent studies, plant-pollinator interactions are also being discovered through metabarcoding of pollen grains and honey (Keller and Steffan-Dewenter 2014). For example, Sickel et al. (2015) identified the source constituents of pooled pollen samples collected from solitary bees using ITS barcodes and Illumina MiSeq, which identified most of the taxa up to species level. Similarly, a parallel study conducted by Richardson et al. (2015), using ITS2, discovered 19 source plant families of pollen grains collected from bee colonies, whereas only 8 families were detected by

microscopic melissopalynology (Erdtman 1943; Kearns and Inouye 1993). Hawkins et al. (2015) used *rbcL* barcoding in combination with 454-pyrosequencing to discern the botanical origin of honey sample and identified a broad range of plant taxa that included 46 plant families and 25 orders. Pornon et al. (2016), using *trnL* and ITS1 metabarcoding technique, identified plant taxa from pollen mixtures, which were 2.5-fold higher than noted through direct field observation. However, such studies focus only on the foraging preferences of pollinators and on the conservation of endangered plant and pollinator species. A recent study of Gous et al. (2015) has provided insights into the foraging behavior of primitive insect pollinators and the change in their floral preferences over time using ITS1, ITS2, and *rbcL* markers. DNA barcoding followed by Illumina next-generation sequencing detected the source plants of pollen collected from ancient bee specimens. Furthermore, all the three plant barcodes successfully generated sequences even for those specimens dating from 1910 and enabled the plant identification up to genus level.

5 Conclusion

The history of plant-animal interaction dates back to the cretaceous period when the angiosperm (flowering) plants evolved first. This interaction is considered as one of the major reasons for the overwhelming diversity of angiosperm plants and their associated animals seen today. In an ecological viewpoint, these interactions represent a major conduit of energy that passes through food cycle to higher trophic levels. Both mutualistic and antagonistic interactions hold equal importance in maintaining the biological diversity stable. Most of the plant-animal interactions are obligatory in nature, and knowledge about the interacting species is a prerequisite to update the current state of the interaction and, if required, to develop strategies to revive the weakening specific interactions. Nevertheless, species level interactions are much complicated to document in the wild. Since most of the interactions are species specific, ecologists might find DNA barcoding technique as a quick and accurate method to unravel the mysterious interactions existing between plants and animals. In addition to this, it offers a noninvasive procedure to quantify feeding habits of animals from fecal matters, which prevents the euthanization of animals. The growing literature demonstrates the great potential of DNA barcoding technique in redefining the interactions of plants with pollinators, frugivores, and herbivores.

Acknowledgments The authors are grateful to Science and Engineering Research Board (SERB), Department of Science and Technology, Government of India, New Delhi, for the financial assistance (SR/S0/AS-84/2012). VRP is grateful to the DST-INSPIRE (IF160266) for the support in the form of research fellowship.

References

Ali MA, Pandey AK, Gyulai G, Park SH, Lee C, Kim SY, Lee J (2015) Introduction to plant DNA barcoding. In: Plant DNA barcoding and phylogenetics, pp 1–14

Bartomeus I, Vilà M, Santamaria L (2008) Contrasting effects of invasive plants in plant-pollinator networks. Oecologia 155:761–770

Boyer S, Benjamin RW, Steve DW (2016) How molecular tools can help understanding species interactions to improve restoration outcomes. Peer J Preprints. https://doi.org/10.7287/peerj.preprints.2020v1

Bruni I, Galimberti A, Caridi L, Scaccabarozzi D, De Mattia F, Casiraghi M, Labra M (2015) A DNA barcoding approach to identify plant species in multiflower honey. Food Chem 170:308–315

Caron GE, Leblanc R (1992) Pollen contamination in a small black spruce seedling seed orchard for three consecutive years. For Ecol Manag 53(1–4):245–261

Chama L, Berens DG, Downs CT, Farwig N (2013) Habitat characteristics of forest fragments determine specialization of plant-frugivore networks in a mosaic forest landscape. PLoS One 8 (1):e54956

Clare EL (2014) Molecular detection of trophic interactions: emerging trends, distinct advantages, significant considerations and conservation applications. Evol Appl 7(9):1144–1157

Corlett RT (1998) Frugivory and seed dispersal by vertebrates in the oriental (Indomalayan) region. Biol Rev Camb Philos Soc 73(4):413–448

Cowan RS, Fay MF (2012) Challenges in the DNA barcoding of plant material. Methods Mol Biol 862:23–33

Darwin CR (1859) On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London

Darwin CR (1862) On the various contrivances by which British and foreign orchids are fertilised by insects. John Murray, London

Davis DD (1962) Mammals of the lowland rain-forest of north Borneo. Bull Singapore Nat Mus 31:3–128

Dunning LT, Savolainen V (2010) Broad-scale amplification of matK for DNA barcoding plants, a technical note. Bot J Linn Soc 164:1–9

Ehrlich PR, Raven PH (1964) Butterflies and plants: a study in coevolution. Evolution 18 (4):586–608

Encinas-Viso F, Revilla TA, Velzen EV, Etienne RS (2014) Frugivores and cheap fruits make fruiting fruitful. J Evol Biol 27(2):313–324

Erdtman G (1943) An introduction to pollen analysis. Chronica Botanica, Waltham, MA

Farrell BD (1998) 'Inordinate fondness' explained: why are there so many beetles? Science 281 (5376):555–559

Fazekas AJ, Kuzmina ML, Newmaster SG, Hollingsworth PM (2012) DNA barcoding methods for land plants. Methods Mol Biol 858:223–252

Futuyma DJ, Agrawal AA (2009) Macroevolution and the biological diversity of plants and herbivores. Proc Natl Acad Sci USA 106(43):18054–18061

Galimberti A, De Mattia F, Bruni I, Scaccabarozzi D, Sandionigi A, Barbuto M, Casiraghi M, Labra M (2014) A DNA barcoding approach to characterize pollen collected by honeybees. PLoS One 9(10):e109363

Galimberti A, Spinelli S, Bruno A, Mezzasalma V, De Mattia F, Cortis P, Labra M (2016) Evaluating the efficacy of restoration plantings through DNA barcoding of frugivorous bird diets. Conserv Biol 30(4):763–773

García-Robledo C, Erickson DL, Staines CL, Erwin TL, Kress WJ (2013) Tropical plant–herbivore networks: reconstructing species interactions using DNA barcodes. PLoS One 8(1):e52967

Giannini N, Kalko E (2004) Trophic structure in a large assemblage of phyllostomid bats in Panama. Oikos 105:209–220

Gonzalez-Varo JP, Arroyo JM, Jordano P (2014) Who dispersed the seeds? The use of DNA barcoding in frugivory and seed dispersal studies. Methods Ecol Evol 5:806–814
Gous A, de Bruin JJ, Willows-Munro S, Eardley C, Swanevelder D (2015) The application of next-generation sequencing barcoding in identifying mixed-pollen samples from a historic bee collection. 6th International Barcode of Life Conference, Guelph, Canada. Genome 58:222
Greenwood MS (1986) Gene exchange in loblolly pine: the relation between pollination mechanism, female receptivity and pollen availability. Am J Bot 73:1443–1451
Gruezo W, Soligam AC (1990) Identity and germination of seeds from feces of the Philippine palm civet (*Paradoxurus philippinensis* Jourdan). Nat Hist Bull Siam Soc 38:69–82
Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S, Frazer KA (2009) Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biol 10(3):R32
Hawkins J, de Vere N, Griffith A, Ford CR, Allainguillaume J, Hegarty MJ, Baillie L, Adams-Groom B (2015) Using DNA metabarcoding to identify the floral composition of honey: a new tool for investigating honey bee foraging preferences. PLoS One 10(8):e0134735
Hayward CE (2013) DNA barcoding expands dietary identification and reveals dietary similarity in Jamaican frugivorous bats. University of Western Ontario – Electronic Thesis and Dissertation Repository. Paper 1697
Hebert PD, Cywinska A, Ball SL, de Waard JR (2003) Biological identifications through DNA barcodes. Proc Biol Sci 270(1512):313–321
Hegland SJ, Nielsen A, Lazaro A, Bjerknes AL, Totland Ø (2009) How does climate warming affect plant-pollinator interactions? Ecol Lett 12(2):184–195
Hilu KW, Liang H (1997) The matK gene: sequence variation and application in plant systematics. Am J Bot 84:830–839
Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. PLoS One 6(5):e19254
Jain SA, De Jesus FT, Marchioro GM, De AraúJo ED (2013) Extraction of DNA from honey and its amplification by PCR for botanical identification. Food Sci Technol (Campinas) 33(4):753–756
Joshi AR, Smith JLD, Cuthbert FJ (1995) Influence of food distribution and predation pressure on spacing behavior in palm civets. J Mammal 76:1205–1212
Jurado-Rivera JA, Vogler AP, Reid CA, Petitpierre E, Gómez-Zurita J (2009) DNA barcoding insect-host plant associations. Proc Biol Sci 276(1657):639–648
Kajtoch Ł (2014) A DNA metabarcoding study of polyphagous beetle dietary diversity: the utility of barcodes and sequencing techniques. Folia Biol (Krakow) 62(3):223–234
Kambli S, Aiswarya S, Manoj K, Varma S, Asha G, Rajesh TP, Sinu PA (2017) Leaf foraging sources of leafcutter bees in a tropical environment: implications for conservation. Apidologie 48(4):473–482. https://doi.org/10.1007/s13592-016-0490-2
Kearns CA, Inouye DW (1993) Techniques for pollination biologists. University Press of Colorado, Boulder, CO
Kearns CA, Inouye DW, Waser NM (1998) Endangered mutualisms: the conservation of plant-pollinator interactions. Annu Rev Ecol Evol Syst 29(1):83–112
Keller A, Steffan-Dewenter I (2014) Pollen DNA barcoding using next generation sequencing. https://doi.org/10.1111/plb.12251
Kitson JJN, Warren BH, Florens FBV, Baider C, Strasberg D, Emerson BC (2013) Molecular characterization of trophic ecology within an island radiation of insect herbivores (Curculionidae: Entiminae: Cratopus). Mol Ecol 22(21):5441–5455
Kowalczyk R, Taberlet P, Coissac E, Valentini A, Miquel C, Kamiński T, Wójcik JM (2011) Influence of management practices on large herbivore diet—Case of European bison in Białowieza Primeval Forest (Poland). For Ecol Manag 261:821–828
Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S (2014) Plant DNA barcoding: from gene to genome. Biol Rev Camb Philos Soc 90(1):157–166
Linhart YB, Busby WH, Beach JH, Feinsinger P (1987) Forager behavior, pollen dispersal, and inbreeding in two species of hummingbird-pollinated plants. Evolution 41:679–682

Lopez JE, Vaughan C (2007) Food niche overlap among neotropical frugivorous bats in Costa Rica. Rev Biol Trop 55(1):301–313
MacIvor JS (2016) DNA barcoding to identify leaf preference of leafcutting bees. R Soc Open Sci 3 (3):150623
Matheson CD, Muller GC, Junnila A, Vernon K, Hausmann A, Miller MA, Greenblatt C, Schlein Y (2008) A PCR method for detection of plant meals from the guts of insects. Org Divers Evol 7:294–303
McClenaghan B, Gibson JF, Shokralla S, Hajibabaei M (2015) Discrimination of grasshopper (Orthoptera: Acrididae) diet and niche overlap using next-generation sequencing of gut contents. Ecol Evol 5(15):3046–3055
Mchatton R (2011) Orchid pollination: exploring a fascinating world. Am Orchid Soc 340–349
Muscarella R, Fleming TH (2007) The role of frugivorous bats in tropical forest succession. Biol Rev Camb Philos Soc 82(4):573–590
Nagarajan M, Nimisha K, Nimisha AK, Sinu PA (2016) The promise of genomics in plant-animal interaction studies. In: Sinu PA, Shivanna KR (eds) Mutualistic interaction between flowering plants and animals. Manipal University Press, Manipal, pp 293–303
Nakashima Y, Inoue E, Inoue-Murayama M, Sukor J (2010) High potential of a disturbance-tolerant frugivore, the common palm civet paradoxurus hermaphroditus (viverridae), as a seed disperser for large-seeded plants. Mammal Study 35(3):209–215
Nishihiro J, Washitani I (1998) Patterns and consequences of self-pollen deposition on stigmas in heterostylous *Persicaria japonica* (Polygonaceae). Am J Bot 85(3):352
Novotny V, Basset Y, Miller SE, Weiblen GD, Bremer B, Cizek L, Drozd P (2002) Low host specificity of herbivorous insects in a tropical forest. Nature 416(6883):841–844
Oberprieler RG, Marvaldi AE, Anderson RS (2007) Weevils, weevils, weevils everywhere. Zootaxa 1668:491–520
Ødegaard F, Diserud OH, Østbye K (2005) The importance of plant relatedness for host utilization among phytophagous insects. Ecol Lett 8:612–617
Ollerton J, Winfree R, Tarrant S (2011) How many flowering plants are pollinated by animals? Oikos 120(3):321–326
Parra V, Vargas CF, Eguiarte LE (1993) Reproductive biology, pollen and seed dispersal, and neighborhood size in the hummingbird-pollinated *Echeveria gibbiflora* (Crassulaceae). Am J Bot 80:153–159
Pegard A, Miquel C, Valentini A, Coissac E, Bouvier F, François D, Taberlet P, Engel E, Pompanon F (2009) Universal DNA-based methods for assessing the diet of grazing livestock and wildlife from feces. J Agric Food Chem 57(13):5700–5706
Pinzón-Navarro S, Jurado-Rivera JA, Gomez-Zurita J, Lyal CHC, Vogler AP (2010) DNA profiling of host–herbivore interactions in tropical forests. Syst Entomol 35(1):18–32
Pompanon F, Deagle BE, Symondson WO, Brown DS, Jarman SN, Taberlet P (2012) Who is eating what: diet assessment using next generation sequencing. Mol Ecol 21(8):1931–1950
Pornon A, Escaravage N, Burrus M, Holota H, Khimoun A, Mariette J, Pellizzari C, Iribar A, Etienne R, Taberlet P, Vidal M, Winterton P, Zinger L, Andalo C (2016) Using metabarcoding to reveal and quantify plant-pollinator interactions. Sci Rep 6:27282
Raye G, Miquel C, Coissac E, Redjadj C, Loison A, Taberlet P (2011) New insights on diet variability revealed by DNA barcoding and high-throughput pyrosequencing: chamois diet in autumn as a case study. Ecol Res 26:265–276
Reddy NP, Bahadur B, Kumar PV (1977) Heterostyly in *Polygonum chinense* L. J Genet 63 (2):79–82
Richardson DM, Allsopp N, D'Antonio CM, Milton SJ, Rejmanek M (2000) Plant invasions – the role of mutualisms. Biol Rev 75:65–93
Richardson RT, Lin CH, Sponsler DB, Quijia JO, Goodell K, Johnson RM (2015) Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. Appl Plant Sci 3(1):1400066

Robledo-Arnuncio JJ, García C (2007) Estimation of the seed dispersal kernel from exact identification of source plants. Mol Ecol 16(23):5098–5109

Robson DB (2014) Identification of plant species for crop pollinator habitat enhancement in the northern prairies. J Pollinat Ecol 14(21):218–234

Sankamethawee W, Pierce AJ, Gale GA, Hardesty BD (2011) Plant-frugivore interactions in an intact tropical forest in north-east Thailand. Integr Zool 6(3):195–212

Sargent RD, Ackerly DD (2007) Plant–pollinator interactions and the assembly of plant communities. Trends Ecol Evol 23(3):123–130

Schallhart N, Tusch MJ, Wallinger C, Staudacher K, Traugott M (2012) Effects of plant identity and diversity on the dietary choice of a soil-living insect herbivore. Ecology 93(12):2650–2657

Scharmann M, Thornham DG, Grafe TU, Federle W (2013) A novel type of nutritional ant–plant interaction: ant partners of carnivorous pitcher plants prevent nutrient export by dipteran pitcher infauna. PLoS One 8(5):e63556

Schiest FP, Johnson SD (2013) Pollinator-mediated evolution of floral signals. Trends Ecol Evol 28 (5):307–315

Schnell IB, Fraser M, Willerslev E, Thomas M, Gilbert P (2010) Characterisation of insect and plant origins using DNA extracted from small volumes of bee honey. Arthropod Plant Interact 4 (2):107–116

Shokralla S, Spall J, Gibson JF, Hajibabaei M (2012) Next-generation DNA sequencing technologies for environmental DNA research. Mol Ecol 21(8):1794–1805

Sickel W, Ankenbrand MJ, Grimmer G, Holzschuh A, Härtel S, Lanzen J, Steffan-Dewenter I, Keller A (2015) Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. BMC Ecol 15:20

Silveira M, Trevelin L, Port-Carvalho M, Godoi S, Mandetta EN, Cruz-Neto AP (2011) Frugivory by phyllostomid bats (Mammalia: Chiroptera) in a restored area in Southeast Brazil. Acta Oecol 37:31–36

Sinu PA, Shibil VK, Manoj K, Rajesh TB (2016) Recent advances on biotic seed dispersal by vertebrates. In: Sinu PA, Shivanna KR (eds) Mutualistic interaction between flowering plants and animals. Manipal University Press, Manipal, pp 221–237

Soininen EM, Valentini A, Coissac E, Miquel C, Gielly L, Brochmann C, Brysting AK, Sønstebø JH, Ims RA, Yoccoz NG, Taberlet P (2009) Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. Front Zool 6(1):16

Srivathsan A, Sha JC, Vogler AP, Meier R (2014) Comparing the effectiveness of metagenomics and metabarcoding for diet analysis of a leaf-feeding monkey (*Pygathrix nemaeus*). Mol Ecol Resour 15(2):250–261

Staudacher K, Wallinger C, Schallhart N, Traugott M (2011) Detecting ingested plant DNA in soil-living insect larvae. Soil Biol Biochem 43(2):346–350

Su S, Sale J (2007) Niche differentiation between common palm civet *Paradoxurus hermaphroditus* and small Indian civet viverricula indica in regeneration degraded forest, Myanmar. Small Carniv Conserv 36:30–34

Teixeira RC, Corrêa CE, Fischer E (2009) Frugivory by *Artibeus jamaicensis* (Phyllostomidae) bats in the Pantanal, Brazil. Stud Neotrop Fauna Environ 44:7–15

Thomas SG, Rehal SM, Varghese A, Davidar P, Potts SG (2009) Social bees and food plant associations in the Nilgiri Biosphere Reserve, India. Trop Ecol 50:79–88

Valentini A, Miquel C, Nawaz MA, Bellemain E, Coissac E, Pompanon F, Gielly L, Cruaud C, Nascetti G, Wincker P, Swenson JE, Taberlet P (2009) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. Mol Ecol Resour 9 (1):51–60

Viana DS, Gangoso L, Bouten W, Figuerola J (2015) Overseas seed dispersal by migratory birds. Proc Biol Sci 283(1822):20152406

Wallinger C, Staudacher K, Schallhart N, Peter E, Dresch P, Juen A, Traugott M (2013) The effect of plant identity and the level of plant decay on molecular gut content analysis in a herbivorous soil insect. Mol Ecol Resour 13(1):75–83

Waser NM, Price MV (1982) A comparison of pollen and fluorescent dye carry-over by natural pollinators of *Ipomopsis aggregata* (Polemoniaceae). Ecology 63:1168–1172

Wilson EE, Sidhu CS, Levan KE, Holway DA (2010) Pollen foraging behaviour of solitary Hawaiian bees revealed through molecular pollen analysis. Mol Ecol 19(21):4823–4829

Wirth R, Meyer ST, Leal IR, Tabarelli M, Lüttge U, Beyschlag W, Murata J (2008) Plant herbivore interactions at the forest edge. Prog Bot 69:423–448

Wong MM, Lim CL, Wilson JJ (2015) DNA barcoding implicates 23 species and four orders as potential pollinators of Chinese knotweed (*Persicaria chinensis*) in Peninsular Malaysia. Bull Entomol Res 105(4):515–520

DNA Barcoding in Forensic Botany

Mohamed Rizk Enan

Abstract Forensic botany is an interdisciplinary area where the knowledge of botany is applied to solve the crime. The ambitious idea of using DNA barcoding for large-scale species identification is already a powerful tool for scientists, and the application of this standard technique appears promising in a range of fields including forensic genetics. Botanical evidence can provide useful leads in forensic investigations; however, many plant materials cannot be identified to the species level only by morphological features. Taking advantage of DNA sequencing and other biomolecular techniques, exact identification of plants becomes useful in selected cases in court producing successful results. Forensic botany can provide significant supporting evidence during criminal investigations. The ubiquitous presence of plant species can be useful in forensics, but the absence of an accurate identification system remains the major obstacle to the present inability to routinely and correctly identify trace botanical evidence. Many plant materials cannot be identified and differentiated to the species level by traditional morphological characteristics when botanical specimens are degraded and lack physical features. By taking advantage of a universal barcode system, DNA sequencing, and other biomolecular techniques used routinely in forensic investigations, two chloroplast DNA regions were evaluated for their use as "barcoding" markers for plant identification in the field of forensics. In this chapter, we review (1) the use of DNA barcoding in forensics, (2) subdisciplines of forensic botany, (3) the history and potential of pollen in forensic studies, (4) DNA barcoding for timber identification, (5) DNA barcoding of pollen grains, and (6) the use of plant forensics in death investigation.

Keywords Forensic · Plant barcoding · Phylogenetic analysis

M. R. Enan (✉)
Molecular Genetics, Biology Department, United Arab Emirates University, Al-Ain, UAE
e-mail: mohamed.enan@uaeu.ac.ae

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_9

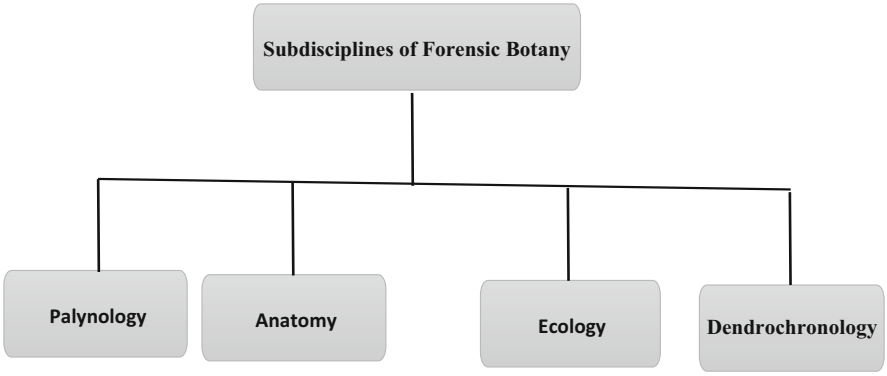


Fig. 1 A schematic diagram of the subdisciplines of forensic botany

1 Introduction

The term forensic comes from the Latin word "*Forensis*" meaning "the forum." Forensic science is often defined as the application of scientific methods and techniques for the purposes of justice. It has been branched into various headings such as forensic biology, forensic anthropology, forensic chemistry, forensic physics, forensic botany, forensic entomology, and forensic serology. Forensic botany is defined as the use of plant evidence in court. In other words, examination of plant materials or remains collected at a crime scene and using them to solve crimes or other legal problems may be loosely termed as forensic botany (Chandra and Sharma 2014). Forensic botany can play a valuable role in criminal investigations, but it has been not used because of the lack of botanical knowledge among most people involved in criminal investigations. Forensic botany is subdivided into several specialties of botanical fields like plant taxonomy, palynology, plant ecology, limnology (Fig. 1), and more recently molecular plant biology. Plant identification can help to determine a sample's geographic origin, provide links between crime scene and individuals, ascertain the possession or trade in forbidden or endangered species, and more (Yoon 1993; Quatrehomme et al. 1997; Coyle et al. 2001, 2005; Mildenhall 2006a, b; Bock and Norris 1997; Dunbar and Murphy 2009). Even if the value of botanical trace evidence in criminal and civil cases has been clearly demonstrated and it is widely accepted as suitable scientific evidence by the courts, it remains ignored by many investigators. Morphology and anatomy rarely confirm the source of the plant, particularly for degraded and fragmented material. Taking advantage of DNA sequencing and other biomolecular techniques, identification of land plants becomes useful in forensics, reducing the expertise necessary in botanical identification and producing successful results in casework application. Analysis of trace botanical evidence is a developing forensic discipline, and techniques based on molecular biology should support and complement traditional forensic botany in the identification of evidence. The problem of reproducibility and standardization, especially from degraded specimens, prevents the use of classical methods in routine

forensic investigations (Korpelainen and Virtanen 2003), but the use of genetic markers could have the potential to overcome shortcomings associated with traditional forensic botany, and it has been recently proposed in forensic applications (Ward et al. 2005; Tsai et al. 2006, 2008). Recently, the use of short standard DNA sequences as a tool for species identification with a standardized protocol, known as DNA barcoding, has been proposed and initiated to facilitate biodiversity and taxonomic studies and enhance forensic analyses (Hebert et al. 2003a; Savolainen et al. 2005; Dawnay et al. 2007). The main purpose of DNA barcoding is to provide rapid and accurate identification of unidentified organisms whose DNA barcodes have already been registered in a DNA database. In order for a region of DNA to be effective as a barcode, it must simultaneously contain enough variability to be informative for identification, be short enough to sequence in a single reaction, and contain conservative regions, which can be used to develop universal primers. A portion of the mitochondrial cytochrome *c* oxidase I gene sequence is currently being used as a universal barcode in certain animal groups and has been proposed also in the forensic field (Hebert et al. 2003b; Dawnay et al. 2007; Ferri et al. 2009).

The forensic botanist requires DNA extraction, amplification, and sequencing. The information generated from plant DNA databases can be used to correlate specific markers with known plant species to establish forensic linkage using sequence similarity and homology search tools (Ferri et al. 2009). DNA barcoding of pollen could transform forensic palynology by making it accessible to a very wide range of forensic laboratories that have available molecular tools and increasing the taxonomic resolution of identification. While this technology has not yet been applied to forensic palynology, recent advances in the development of standardized DNA barcoding loci for plants (CBOL Plant Working Group 2009), as well as improvements in sequencing technologies for processing samples with multiple species, indicate that the tools are now in place to apply this transformative technology (Valentini et al. 2010; Kraaijeveld et al. 2015; Keller et al. 2015; Richardson et al. 2015).

2 DNA Barcodes in Plant Forensics

Based on the existing literature, there is currently no available universally usable region, and it is generally agreed that a multilocus approach based on chloroplast data is the most effective strategy for species identification and recognition in plants (Kress et al. 2005; Cowan et al. 2006; Chase et al. 2007, Hollingsworth 2008). A variety of loci have been recently suggested as potential DNA barcodes in plants, in both the nuclear and plastid genomes (Kress and Erickson 2007; Sass et al. 2007; Fazekas et al. 2008). Genetic methods are routinely used in forensics for the identification of human individuals, usually with relatively inexpensive, high-throughput, non-sequencing-based PCR methods, such as short tandem repeats (STRs) or single nucleotide polymorphisms (SNPs) (Lee et al. 1994; Linacre and Graham 2002; Børsting and Morling 2015). Genetic methods are also common in the

forensic analysis of targeted species, such as for wildlife (Ogden and Linacre 2015), drug trafficking (Mello et al. 2016), and food adulteration (Galimberti et al. 2012); Bosmali et al. 2012). DNA barcoding is not a new science; it has been used with animal species based on the mitochondrial cytochrome *c* oxidase I (COX1) (Newmaster et al. 2006). But the same DNA sequence cannot be used for plant barcoding because plants do not have the same amounts of mtDNA available as animals. Therefore, it becomes more challenging (Newmaster et al. 2006). To make it a lot easier to identify plants when dealing with forensic casework and trying to work on criminal investigations, having a database of plant DNA that can be used as comparative research is greatly helpful. So the scientific community has come up with DNA barcodes, created from short standard DNA sequences to use for species identification, to aid in studies and help with forensic analysis (Ferri et al. 2009). Although for DNA to be viable to become a DNA barcode, it has to have enough information for it to be identified, so if someone found an unknown species at a crime scene and hoped to create a DNA barcode for it, that might not be possible if the sample is not large enough. So even though the DNA barcoding can be extremely helpful in establishing a database, it can also be tedious work that might not even give results. There is not yet a universal region used on the DNA; there are still many loci being used to identify plants, and that is the most effective, although there have been several loci that have been suggested for universal DNA barcodes in plants (Ferri et al. 2009). Two main benefits of using DNA barcoding in forensics are:

1. The same method is used across multiple taxonomic groups, so it allows species to be identified without having to involve narrowly defined taxonomic experts.
2. It allows taxonomic identification of parts of organisms that do not display diagnostic morphological characters.

Combining DNA barcoding with high-throughput sequencing allows forensic analysis of complex mixtures. Methods for using DNA barcoding to associate an item with a particular location have recently been developed based on soil DNA (Young et al. 2014; Giampaoli et al. 2014) and plant material (Ferri et al. 2015). Forensic scientists are always looking for new methods to use DNA to help investigators solve crimes and to make sense of evidence. The barcoding and DNA databases that are being implemented in forensic botany will help to speed up routine investigations of plant evidence and detection of new species and aid in the forensic analysis of other, unknown plant artifacts. Within the forensic perspective, it should be taken in consideration that not all techniques meet forensic standards, but we should also consider that the Barcoding Project can warranty the level of reproducibility and standardization necessary to present evidences in a court.

3 Main Subdisciplines of Forensic Botany

3.1 Forensic Ecology and Limnology

Ecology is the branch of biology that deals with the relations of organisms to one another and to their physical surroundings. When we take plant ecology, it includes the distribution abundance of plants in a region, their effects upon environmental factor, and interactions among and between plants and other organisms. Knowledge of ecology is useful for the forensic investigators in determining the relation of victims, crime scene, and suspects, as a forensic ecologist, one needs to know several things in relation to ecology, not only the structure of plants but also their natural and seminatural habitat, biotic and abiotic component of the habitat, how one species affects others, their possible and probable time of abundance, growth, structure, and overall functions. Apart from these, forensic ecologists are expected to know about garden and gardening too. They are required to keep the knowledge of plantation parks, ponds, canals, and wastelands. Forensic ecologists are not expected to be an expert in every aspect of ecological sciences, knowledge of vegetation, soil sediments, and aquatic and terrestrial habitats which are some essential areas for investigations (Whiltshire 2009). Individual may develop an interest and subsequent expertise in any specific area. The study of the structural and functional interrelationships of organisms of inland waters as they are affected by their dynamic physical, chemical, and biotic environments is called limnology. Limnology involves the study of both freshwater and saline inland water; freshwater bodies are of more importance to human activities. Thus forensic limnology is more concerned to freshwater ecology. Various cases exemplify the applicability of plant ecology to forensic investigation. In one case, victim was brutally attacked to death and was drowned into a nearby pond. The accompanied person with victim managed to escape and informed to the immediate possible option. Investigators resumed water-soaked, mud-encrusted footwear and clothing from the apprehended suspect. Investigators also obtained samples of pond water sediments and aquatic vegetation from the pond aiming to establish the linkage between the clothing of suspects, victims, and pond biota. Investigators with the basic ecological application and certain limnological tests established the linkage. Numerous species of diatoms and planktonic algae were recovered from the victims' and suspects' clothing and reference samples of pond sediments (Siver et al. 1994). Quantitative diatom based reconstruction technique to confirm drowning as cause of death and localize the site of drowning has to be developed. They collected samples from the local/regional water bodies to act as control in the examination of diatom assemblages associated with lungs and clothing of the subject (Horton et al. 2006).

3.2 Forensic Anatomy

Plant anatomical features have long been used by archeologists and paleontologists to characterize archeological sites and since the 1930s have become increasingly more common in forensic applications (Lane et al. 1990). The cell wall is particularly important for two reasons: it is not easily digested by most organisms and therefore persists when other plant features are destroyed (Bock and Norris 1997), and the size, shape, and pattern of cell walls are often taxon-specific (Lane et al. 1990). Lancia and Conforti (2013) studied the growth rate of the bryophyte *Leptodyctium riparium*, which was used in estimating the postmortem interval of some human skeletal remains that were found in a wooded area near Perugia, in Central Italy. The construction of plant's body and its components can provide a wide variety of forensic evidences. The rigid external cell wall covering the plant cells, unlike the animal cells, does not allow the plant cells to wash away or degenerate easily. The varying percentage of cellulose and lignin making up the cell wall are not digestible by animals except ruminants, termites, and a few other arthropods as far as cellulose is concerned, whereas with lignin it is somewhat more difficult to digest due to its being more complex structurally and chemically which can be broken down only by a few fungi. In addition to this, cells that make up wood have a secondary wall that has much higher percentage of lignin, which adds on to the strength and long life of these cells. In forensic science, the identification of wood is done on the basis of the physical and anatomical properties of wood. Physical properties include color, odor, weight, hardness, and texture of the wood sample. Also, the water extract of some timbers shows distinct fluorescence which helps in identifying the unknown wood samples (Chandra and Sharma 2014). Anatomical properties include whether the pores are present or absent, and if present then the further identifications are done on determining the pore number, pore size, and pore arrangement. The manner in which cell walls are laid down and the patterns formed by cells are peculiar to given taxonomic groups. The nonwoody (herbaceous) plants have characteristic cell shapes, arrangements of cells of the epidermis, and combinations of cell types in their pulpy parts. Some cells contain crystals of calcium oxalate or starch grains of types that vary among species. Identification of starch grains is based on three characteristics—size, shape, and presence or absence as well as shape of markings on the starch grains called the hilum. Identification of starch grain is easy on freshly prepared starch. In old starch or boiled starch or chemically modified starch, the hilum changes from dot to transverse or transverse to satellite and satellite to ruptured grain. Thus, too much reliance should not be placed on hilum unless the history of starch is known. Comparison of unknown sample with known sample is always desirable. In case of decomposed starches, identification should be based on few grains which are left intact (Sharma 2014). Correct identification of plant samples depends on trained botanists' abilities to gain access to recorded information on the characteristics of the species to which the sample belongs. Therefore, anyone who identifies a new sample of plants must publish the contents as all information is needed and can help in solving the criminal cases.

3.3 Forensic Palynology

Forensic palynology, a subdiscipline of forensic botany, utilizes pollen grains as trace evidence for criminal investigations. First used in the 1950s, pollen evidence is analyzed for forensic purposes prevailingly in countries such as Australia, New Zealand, and the UK (Bryant and Jones 2006). Pollen may be present as a fine or coarse powder and consists of the microgametophytes of seed plants (which produce the male gametes). An individual tree can produce thousands to millions of grains in a single pollinating season. Wind-pollinated plants produce an abundance of pollen dispersed over long distances, while insect-pollinated plants produce smaller amounts dispersed over shorter distances. Due to its strong cellular wall, pollen deposited can persist for months to years after initial dispersal, even after the original plant is not present anymore (Mildenhall et al. 2006). Therefore, the distribution and variation of pollen populations in two different locations only a few miles apart may be considerably different (Horrocks 2004). Moreover, the seasonality of pollen can be studied with the help of pollen calendars (Mercuri 2015). A study of Montali et al. (2006) demonstrated that pollen calendars can be used as reference tools for determining the season of death in case of murders or other crimes. The persistence of small trace particulates on clothing is of forensic importance and is dependent on multiple factors, including size, texture, and the type of fabric. It is identified that following particulate transfer, the time and degree of activity of the subject (host) until recovery are crucial factors that are dependent on one another. While many studies describe the persistence and transfer of fibers and other trace evidence on clothing (Pounds and Smalldon 1975a, b, c; Hicks et al. 1996; Allen and Scranage 1998; Petterd et al. 1999), few have investigated pollen grains (Horrocks et al. 1999; Bull et al. 2006; Zavada et al. 2007; Morgan et al. 2014a, b). Zavada et al. (2007) demonstrated that "clothing acts as a quantitative and qualitative collector, and produces a pollen profile that exhibit regional seasonal trends similar to patterns generated using volumetric collectors." Traditionally, pollen is identified using microscopy techniques based on morphological characteristics. Morgan et al. (2014a, b) and Mildenhall (2006a, b) described that, palynologists use an "assemblage" consisting of hundreds of grains to determine the types and percentage of the various pollen species present. However, in forensic cases, hundreds of grains are not always available, and individualization of pollen assemblages based on morphology alone can be very difficult. Male sex cells (containing DNA) are trapped inside a rigid pollen grain wall and protected until the grain is crushed. In the case of *Pinus*, one pollen grain can include many cells; however, the sperm nuclei may develop with or without cells. Currently, the application of molecular biology techniques to pollen grains is being used in other scientific areas such as environmental sciences, ecology, and agriculture (Zhou et al. 2007; Matsuki et al. 2008). A DNA fragment found on the *rpoB* region of chloroplast DNA has been used to genotype *Pinus* pollen grains. The internal transcribed spacer (ITS) has been used to identify *Chenopodiaceae* pollen grains in surface soil down to the species level (Zhou et al. 2007). Short tandem repeat (STR) analysis has been used to

monitor the dispersion of single pollen grains from one area to another by insects (Matsuki et al. 2008). Forensic palynology is the use of pollen to link persons or objects with particular places and times (Horrocks and Walsh 1998; Taylor and Skene 2003; Bryant and Jones 2006; Mathewes 2006). This technique is of great utility to forensics because:

1. Pollen is a nearly ubiquitous feature of the environment.
2. Different geographic locations have different pollen signatures, allowing for inference related to spatial tracking.
3. Plants flower at different times, allowing for temporal inference.
4. Pollen is extremely durable (hence its widespread use in paleontological studies) and thus can be utilized for forensic studies for decades or longer after sample collection (Horrocks and Walsh 1998; Bryant and Jones 2006; Mathewes 2006; Mildenhall 2006a, b; Mildenhall et al. 2006; Wiltshire 2006; Walsh and Horrocks 2008).

Like many other palynological applications, forensic palynology as currently practiced is reliant on visual microscopic identification of pollen grains by an expert palynologist (Bryant and Jones 2006; Mildenhall et al. 2006; Walsh and Horrocks 2008; Arguelles et al. 2015), and DNA metabarcoding could very likely increase its applicability to a broader range of situations (Bell et al. 2016). For example, forensic palynology could take greater advantage of DNA barcoding when combined with a universal database of geographic and temporal knowledge of plants. Such a database could also be of potential utility to fields outside of forensics (airborne pollen monitoring for allergens, pollination biology, biodiversity inventories, and potentially even monitoring of plant populations). As with any genetic analysis, DNA barcoding requires destructive sampling of pollen grains, which means the pollen can no longer be analyzed with morphological methods. One possible work-around to this issue is to split pollen samples into partitions for DNA extraction, morphological examination, and permanent storage. Comparing the results of the DNA barcoding and morphological identifications for consistency could validate the accuracy of this approach. Although work is needed on these technical issues, and others including in some cases adaptations for very small samples, the method is very close to being feasible for routine analysis in forensics, and improvements are occurring rapidly. Forensic palynology has been particularly useful in cases where evidence points to transfer of objects between geographic locations or where a crime has occurred in a location with a distinct species composition. For example, following the Bosnian war, mass graves were uncovered, including primary sites (where bodies were initially buried) and secondary sites (where bodies were moved and reburied at a later date). By identifying pollen species and soil composition, the bodies could be matched to their primary sites (Brown 2006). Reviews of the use of forensic palynology (Bryant and Jones 2006; Mathewes 2006; Montali et al. 2006) discuss the potential of the technique but recognize that its use is currently limited. As mentioned previously, three factors drive this limitation: the small pool of expert palynologists who can accurately identify pollen to the species level, the time-consuming nature of microscopic identifications, and the limited taxonomic

resolution of pollen determinations in many cases. While all of these limitations are important, the last point (taxonomic resolution) is a particularly critical one.

Forensically, the pollen grains which are commonly found hold less evidentiary value than those which are uncommon or belong to the species which is poorly distributed. In a case reported in Auckland, New Zealand, a woman alleged that the suspect had raped her in an alleyway few meters away from his car, while the suspect claimed that he never moved away from his car and nor did he entered the mentioned alleyway. Adding to this he strongly pointed out that he had never had any sexual intercourse with the victim and the soil found on his clothes is from the driveway area. During the investigation the officials collected the soil samples from both the mentioned areas and also from the suspects' clothes. The analysis of soil samples which also contain pollen grains proved the sexual assault taking place in alleyway; the suspect was convicted later on the basis of this evidence.

3.3.1 The Basic Constituents for Forensic Palynology

Given its advantages over traditional microscopic pollen identification, it may seem surprising that DNA barcoding of pollen has not already been incorporated into forensic palynology, security intelligence, or many other fields to which it could be applied. The key reason for this lack of uptake is that the combination of technologies needed for successful DNA barcoding of mixed pollen samples has only been recognized in the last years (Valentini et al. 2010; Kraaijeveld et al. 2015; Keller et al. 2015; Richardson et al. 2015). In order to successfully conduct DNA barcoding of pollen from forensic samples, three elements are needed: first, a set of genetic markers that can be amplified across the seed plants (those that produce pollen) and that can reliably distinguish between a high proportion of species when sequenced; second, a DNA isolation and sequencing method appropriate for the context of the investigation; in many cases this would include a "next-generation" or "high-throughput" sequencing (HTS) method that generates multiple reads per sample of the set of DNA barcoding markers, allowing the simultaneous identification of several species from a single mixed-species forensic sample; and third, a database containing DNA sequences of the genetic markers for the majority of species of seed plants. The first component, a consensus genetic marker set for DNA barcoding of plants, was published at the end of 2009 (Hollingsworth et al. 2009). A recent study recommended a similar marker set for forensic botany (Ferri et al. 2015), and methods have been developed for single-species pollen DNA barcoding, based on the standard plant DNA barcoding loci, regions of the plastid DNA genes *rbcL* and *matK* (Bell et al. 2016). In terms of the second component, standard protocols have been developed for pollen DNA isolation, and these are now available in commercial kits. Sanger sequencing methods have not been standardized for pollen analysis, and depending on the situation, a HTS method is more applicable. The use of HTS platforms in forensic science has been reviewed elsewhere (Børsting and Morling 2015). HTS sequencing platforms have existed commercially since the mid-2000s (Shendure and Ji 2008), but the key limitations to combining DNA barcoding with

HTS (i.e., "metabarcoding" (Taberlet et al. 2012) include significant error rates and insufficient read lengths relative to the standard plant DNA barcoding loci (partial *rbc*L and *mat*K sequences) which are approximately 500–600 base pairs (bp) in length, although "minibarcodes" of 200–300 bp in length can identify species at least to family level (Little 2014). Until recently, the only sequencing platform able to provide reads of this length was from 454 Life Sciences (Roche Diagnostics Corporation, Branford, USA), and this platform has been successfully used for metabarcoding of plant material (Valentini et al. 2009; Pompanon et al. 2012; Shokralla et al. 2014), including pollen. This platform, however, is no longer in production, and will not be supported beyond mid-2016. More recently, the Ion Torrent platform (Life Technologies, Thermo Fisher Scientific, Waltham, USA) has been used for the sequencing of mixed-species pollen samples, but the application of the method has been limited to short (200 bp) sequencing reads of a noncoding, chloroplast marker (the *trnL* gene) that has not been widely used as a core plant DNA barcode. The third component for DNA barcoding, a database of taxonomic reference sequences, is currently the most limiting. Reference databases such as the Barcode of Life Data Systems (BOLD: http://boldsystems.org) and the International Nucleotide Sequence Database Collaboration (http://www.ncbi.nlm.nih.gov/genbank/collab), which comprises GenBank, the DNA Data Bank of Japan (DDBJ), and the European Molecular Biology Laboratory (EMBL), currently only include barcode sequences for a small proportion of global plant biodiversity. On September 21, 2015, there were 62,957 seed plant species with sequences on BOLD and 113,741 seed plant *rbcL* sequences (including multiple sequences for many species) on NCBI. This compares to an estimated total of 450,000 flowering plant species on earth (Pimm and Joppa 2015). This level of coverage varies between geographic locales, with some regions more comprehensively analyzed than others. It is worth noting that similar knowledge gaps also exist in traditional palynological identification via microscopy, for which reference collections to identify grains to species are also limited (Wiltshire 2006). These reference collections are usually found in the personal collections of experts, and may not be readily accessible (Bryant and Jones 2006). While the lack of geographic and taxonomic coverage of plant species in sequence databases limits the ability to sequence unknown samples, the coverage in the existing databases continues to increase over time, especially following the relatively recent standardization of loci for plant DNA barcoding (CBOL Plant Working Group 2009).

3.4 Forensic Plant Biotechnology

Plant biotechnology in recent few decades has been developed as an exciting area of plant sciences. Through this technology various biological manipulation and interpretation can be done. Techniques like DNA fingerprinting, DNA barcoding, and molecular markers are extensively used these days by investigators to solve the cases. Unlike the conventional taxonomic identification, the morphological character

is limited in forensic investigation, plant species found as evidence material are in fragment, and in such instances identification of the species with the conventional available key in any flora becomes difficult. Though DNA barcoding is one option, it has limited scope in plant science as there is no gene in the plant which can be the barcode unlike in animal where the mitochondrial cytochrome *c oxidase* I has become universal barcode. In these circumstances one has to depend on the biotechnological tools such as "short tandem repeats (STRs)," "single-nucleotide polymorphisms (SNPs)," "simple sequence repeats (SSRs)," and variable number tandem repeats (VNTRs) (Zaya and Mary 2012). Eurlings et al. (2013) have reported forensic identification of *Rauvolfia* sp. using DNA barcoding technique which can be used to control illegal international trade as manually the root samples cannot be identified at species level. Gilmore et al. (2003) diagnosed the profile of 93 cannabis plants (*Cannabis sativa*) using STRs. Bryophytes can be the wise choice for forensic investigation in a case where the criminals walk in a forest or semi-urban area and bryophyte fragments get attached to shoes even after a wearer walks for several hours. Also DNA of bryophytes stays intact in fluctuating environment which allows DNA profiling for evidence (Virtanen et al. 2007). Therefore, one can apply numerous tools of biotechnology in multidimensional forensic analysis and evidence.

4 DNA Barcoding for Timber Identification

DNA barcoding used to identify the species of an individual based on variation at specific gene regions (Hebert et al. 2003c). In plants, two chloroplast gene regions maturase K (*mat*K) and ribulose-bisphosphate carboxylase (*rbc*L) are currently used as standard, but can only distinguish ~70% of plants, and usually require analysis of additional local barcode regions for species-level identification (CBOL Plant Working Group 2009). The potential application of DNA barcoding for timber identification has been demonstrated using the internal transcribed spacer (ITS) gene region in the mahogany family (Muellner et al. 2011). One of the greatest challenges of DNA barcoding in timber is that DNA extracted from wood is generally of poor quality, and it is often not possible to sequence the large fragments associated with the standard barcoding regions, meaning that shorter informative regions need to be developed to attain successful identification via DNA barcoding. DNA barcoding has been criticized in the past for seeking to circumvent the need for basic taxonomy when used for species discovery; however, its use in specimen identification is widely accepted (Will et al. 2005). Identification outcomes via DNA barcoding are more limited when applied to taxonomically understudied clades containing closely related species (Meyer and Paulay 2005) and are likely due to problems associated with accurately determining the levels of sequence variation within and between species. The utility of DNA barcoding for forensic identification purposes is gaining recognition and is expected to continue to rise in popularity as capabilities increase and costs of sequencing come down (Iyengar 2014; Linacre and Tobe 2011). The

existence of extensive online sequence databases adds to the future utility of this method for forensic timber identification, such as the Barcode of Life initiative (Ratnasingham and Hebert 2007) whose BOLD database at the time of writing contains over 58,000 species with barcode sequences from the conifers and angiosperms.

5 Dendrochronology in Forensic Timber Identification

Dendrochronology is the study of tree rings. These annual growth rings occur because the xylem cells get gradually smaller in radius as the growth season proceeds into the dormant season. There is a rapid change in size from small, late season cells to the large, early season cells of the following spring (Willey and Heilman 1987). The number of rings gives the tree's age, and the variation in ring width reflects environmental conditions. Ring patterns of several samples from a given geographic area subject to similar environmental conditions are cross-dated, giving standardized chronologies (curves) for different species in different areas, to which specimens of unknown origin can be compared. Dendrochronology is an important technique in a number of disciplines, including archeology, paleontology, geomorphology, climatology, and ecology. Forensic applications concern the dating of wooden objects and matching objects with crime scenes using the wood's morphological features. The science of dendrochronology is another visual technique with the potential for use in forensic timber identification. Dendrochronology is typically applied to elucidate past climates but also has the potential to provide an age and provenance of trees. The ability to correctly assign provenance using dendrochronology is probably limited, although it has been successfully used to identify the origins of archeologically important timbers (Haneca et al. 2005). The approximate felling date of a tree can potentially be determined if timber products possess the outermost growth rings and bark (Wolodarsky-Franke and Lara 2005; Yaman and Akkemik 2009), but error margins can be substantial and therefore problematic (Jones and Daniels 2012). Individual identification is also a possibility, as growth rings can be "matched" where they line up between pieces of wood from the same individual, although how consistent these patterns are across the entire trunk length of a tree is as yet unclear. Visual dendrochronological analyses are occasionally applied to forensic timber identification. However, given that the majority of illegally logged timber originates from tropical areas, where distinct growth rings are rare, dendrochronology is likely to have limited application globally.

6 Examination of Plant Trace Evidence

Plants are everywhere and, by default, so is botanical trace evidence. Because plants and plant evidence are commonly found, it can be exceedingly helpful in criminal investigations. Although, like many other fields of forensic science, botanical forensics is not particularly widely used, there is not a large group of people who are knowledgeable about it. By using forensic botany to identify plants that are found in crime scenes can have a profound effect on the outcome of a case to test the possession of illegal species and prove the geographic source of the plant (Ferri et al. 2009). Since plants are everywhere and there are so many species of plants, scientists needed to come up with a way to make them easier to identify and use, even for people not specialized in forensic botany. So by "taking advantage of DNA sequencing and other bimolecular techniques, identification of land plants becomes useful in forensics, reducing the expertise necessary in botanical identification and producing successful results in casework application" (Ferri et al. 2009). As with every other piece of evidence that a forensic scientist examines, it can only be beneficiary to the outcome of the investigation if the crime scene investigator collects all necessary pieces to be examined. If a forensic scientist specializes in testing botanical evidence for DNA to determine the origin of the plant and possibly if a corpse had been transported long distances, then that scientist would need to be given those pieces of botanical evidence to figure that out. So CSIs need to realize that nothing can be overlooked at a crime scene because it is almost impossible to know which piece of evidence might contain the key piece of DNA to solve the crime. An article in the *Croatian Medical Journal* stresses the importance of proper plant evidence collection and preservation if someone is planning to use that to aid in a death investigation. To a CSI, a large branch that is obviously out of place found near a victim would likely be collected as evidence, but as far as grains of pollen, it would not be as obvious to collect them, and they would only be collected if the investigator is already aware of the potential of there being such evidence; and oftentimes extensive training in botany is needed before an investigator would know which plants are indigenous and not (Coyle et al. 2005). Once all evidence is collected from a crime scene, it must be carefully transported back to a lab for further examination (Fig. 2). As with all other forms of evidence, plant evidence can be contaminated; and if the goal is to DNA type that specific piece of plant evidence, if it is contaminated, then the results may be false. Oftentimes at crime scenes if a plant looks out of place, they do a DNA test to try and figure out the species of the plant to see if the plant is indigenous to the area or not. Then if the DNA test proves that the species is out of place, then that can give leads about the crime that occurred. Although when the plan is to do a DNA test of a piece of plant evidence, then there needs to be multiple samples from more than one plant of the same species so that the data is valuable enough for comparison and statistical analysis (Coyle et al. 2005). So it is going to be difficult to test plant evidence at a crime scene when it appears to be out of place because there might not be anything to make a comparison. But it

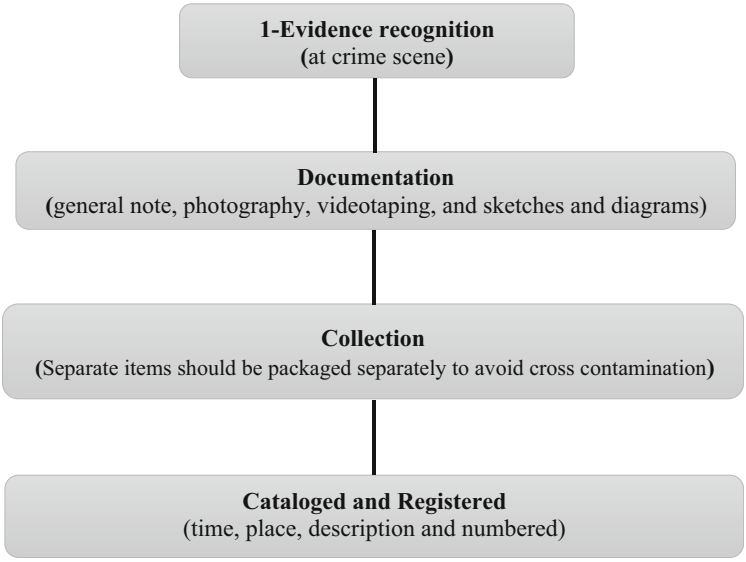

1-Evidence recognition
(at crime scene)
Documentation
(general note, photography, videotaping, and sketches and diagrams)
Collection
(Separate items should be packaged separately to avoid cross contamination)
Cataloged and Registered
(time, place, description and numbered)

Fig. 2 A schematic diagram of the appropriate sequence for forensic evidence

would certainly not be impossible, because upon further investigation of the area, there may be more clues that all come back to that plant.

7 The Use of Plant Forensics in Death Investigation

There are many DNA plant databases that are created to help keep all the species of plants organized and to make the DNA profiles easier to access for scientists testing plant evidence. These databases also help to outline which species of plants belong to certain areas of the world and which ones would be out of place if found there. Coincidentally, the databases also help with death investigations that involve plant evidence. There is a story about a death of a young woman that happened in Taiwan, and when they found her body, they were not sure if it was a homicide or a suicide. After a forensics team investigated the body, they noticed that there were stems and berries in her hair that were clearly not indigenous to the area her body was found (Coyle et al. 2005). When the investigators did some searching around the area where her body was found, they noticed that the evidence in her hair matched the plants that were on one of the window ledges of the building by the scene (Coyle et al. 2005). Then after further investigation of all of the evidence and the autopsy, they were able to conclude that she had killed herself by jumping and hit her head on the window ledge where the plants were (Coyle et al. 2005). So in this case, the botanical evidence ultimately solved the case, and because of the plants, forensics knowledge of the investigators, and possible DNA plant databases, it was a lot

easier. Forensic botany is useful in the examination of gastric contents in homicide victims. It can help determine the contents and location of a victim's last meal. If a victim has been moved from the location of death, plant analysis can determine whether plant remnants on the body are characteristic of the vegetation where the body was found or of some other location where the victim was killed (Chandra and Sharma 2014). The abundant presence of plant species makes botanical trace evidence useful for many aspects of criminal investigations but remains underused in forensics due to a lack of experience and botanical knowledge among investigators as well as the inability to routinely identify degraded specimens by morphology.

8 Summary

Botanical evidence is useful in criminal cases, but it is still a much underutilized resources. Plant evidence is commonly associated with crime scenes and victims and suspects. Typically, stems, leaves, seeds, and flowers are useful as trace botanical evidences; also microscopic pollen may be used to demonstrate an associate link between scenes and evidentiary items. We predict that DNA barcoding will transform forensic palynology, greatly expanding the use of pollen as a biomarker, and, more importantly, will give forensic scientists new leads and evidence toward enhancing global security and justice (Bell et al. 2016). Forensic botany, though, not new but has emerged as an important discipline in solving the mysteries and crimes. It is an interdisciplinary of various branches of botany where a forensic botanist has to acquire knowledge of different fields of botany. With recent advancement in molecular technology, investigation with application of forensic botany has become more precise and accurate. A full-proof and reproducible techniques specific to region and cases have to be developed. The use of plant-based public database and its development according to the need of forensic analyst will help the investigating officer for easy and comparatively fast closure of the case.

References

Allen TJ, Scranage JK (1998) The transfer of glass—Part 1: transfer of glass to individuals at different distances. Forensic Sci Int 93:167–174

Arguelles P, Reinhard K, Shin DH (2015) Forensic palynological analysis of intestinal contents of a Korean mummy. Anat Rec (Hoboken) 298:1182–1190

Bell KL, Burgess KS, Brosi BJ (2016) DNA barcoding of pollen. In: de Vere N (ed) Plant DNA barcoding methods and protocols. Springer, New York

Bock JH, Norris DO (1997) Forensic botany: an under-utilized resource. J Forensic Sci 42:364–367

Borsting C, Morling N (2015) Next generation sequencing and its applications in forensic genetics. Forensic Sci Int Genet 18:78–89. https://doi.org/10.1016/j.fsigen.2015.02.002

Bosmali I, Ganopoulos I, Madesis P, Tsaftaris A (2012) Microsatellite and DNA-barcode regions typing combined with High Resolution Melting (HRM) analysis for food forensic uses: a case study on lentils (*Lens culinaris*). Food Res Int 46:141–147

Brown AG (2006) The use of forensic botany and geology in war crimes investigations in NE Bosnia. Forensic Sci Int 163:204–210
Bryant VM, Jones GD (2006) Forensic palynology: current status of a rarely used technique in the United States of America. Forensic Sci Int 163:183–197
Bull PA, Morgan RM, Sagovsky A, Hughes GJA (2006) The transfer and persistence of trace particulates: experimental studies using clothing fabrics. Sci Justice 46:185–195
CBOL Plant Working Group (2009) A DNA barcode for land plants. Proc Natl Acad Sci USA 106:12794–12797
Chandra R, Sharma V (2014) Forensic botany: an emerging discipline of plant sciences. Indian Botanists Blog-o-Journal. http://www.indianbotanists.com/2014/03/forensic-botany-emerging-discipline-of.html
Chase MW, Cowan RS, Hollingsworth PM et al (2007) A proposal for a standardized protocol to barcode all land plants. Taxon 56:295–299
Cowan RS, Chase MW, Kress JW, Savolainen V (2006) 300,000 species to identify: problems, progress and prospects in DNA barcoding of land plants. Taxon 55:611–616
Coyle HM, Ladd C, Palmbach T, Lee HC (2001) The green revolution: botanical contributions to forensic and drug enforcement. Croat Med J 42:340–345
Coyle HM, Lee CL, Lee HC, Lin WY, Palmbach TM (2005) Forensic botany: using plant evidence to aid in forensic death investigation. Croat Med J 46:606–612
Dawnay N, Ogden R, McEwing R, Carvalho GR, Thorpe RS (2007) Validation of the barcoding gene COI for use in forensic genetic species identification. Forensic Sci Int 173:1–6
Dunbar M, Murphy TM (2009) DNA analysis of natural fiber rope. J Forensic Sci 54:1–6
Eurlings M, Lens F, Pakusza C, Peelen T, Weiringa JJ, Gravendee B (2013) Forensic identification of Indian Snake Root (*Rauvolfia serpentina* Benth. Ex kurz) using DNA barcoding. J Forensic Sci 58:822–830
Fazekas AJ, Burgess KS, Kesanakurti PR et al (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS One 3:e2802
Ferri G, Alù M, Corradini B, Licata M, Beduschi G (2009) Species identification through DNA "barcodes". Genet Test Mol Biomarkers 13:421–426
Ferri G, Corradini B, Ferrari F, Santunione AL, Palazzoli F, Alu' M (2015) Forensic botany II, DNA barcode for land plants: which markers after the international agreement? Forensic Sci Int Genet 15:131–136. https://doi.org/10.1016/j.fsigen.2014.10.005
Galimberti A, De Mattia F, Losa A, Bruni I, Federici S, Casiraghi M et al (2012) DNA barcoding as a new tool for food traceability. Food Res Int 50:55–63
Giampaoli S, Berti A, Di Maggio RM, Pilli E, Valentini A, Valeriani F et al (2014) The environmental biological signature: NGS profiling for forensic comparison of soils. Forensic Sci Int 240:41–47
Gilmore S, Peakall R, Robertson J (2003) Short tandem repeat (STR) DNA markers are hyper variable and informative in *Cannabis sativa*: implication for forensic investigations. Forensic Sci Int 131:65–74
Haneca K, Wazny T, Van Acker J, Beeckman H (2005) Provenancing Baltic timber from art historical objects: success and limitations. J Archaeol Sci 32:261–271
Hebert PDN, Cywinska A, Ball SL, de Waard JR (2003a) Biological identifications through DNA barcodes. Proc R Soc Lond B 270:313–321
Hebert PDN, de Waard JR, Ratnasingham S (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc R Soc Lond B 270:S96–S99
Hebert PDN, Ratnasingham S, De Waard JR (2003c) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc Biol Sci 270:S96–S99
Hicks T, Vanina R, Margot P (1996) Transfer and persistence of glass fragments on garments. Sci Justice 36:101–107
Hollingsworth PM (2008) DNA barcoding plants in biodiversity hot spots progress and outstanding questions. Heredity 101(1):1–2. https://doi.org/10.1038/hdy.2008.16

Hollingsworth ML, Clark AA, Forrest LL, Richardson J, Pennington RT, Long DG et al (2009) Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. Mol Ecol Resour 9:439–457
Horrocks M (2004) Sub-sampling and preparing forensic samples for pollen analysis. J Forensic Sci 49:1–4
Horrocks M, Walsh KAJ (1998) Forensic palynology: assessing the value of the evidence. Rev Palaeobot Palynol 103:69–74
Horrocks M, Coulson SA, Walsh KAJ (1999) Forensic palynology: variation in the pollen content of soil on shoes and in shoeprints in soil. J Forensic Sci 44:119–122
Horton BP, Boreham S, Hillier C (2006) The development and application of a diatom-based quantitative reconstruction technique in forensic science. J Forensic Sci 51:643–650
Iyengar A (2014) Forensic DNA analysis for animal protection and biodiversity conservation: a review. J Nat Conserv 22:195–205
Jones EL, Daniels LD (2012) Assessment of dendrochronological year-of-death estimates using permanent sample plot data. Tree Ring Res 68:3–16
Keller A, Danner N, Grimmer G, Ankenbrand M, von der Ohe K, von der Ohe W et al (2015) Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. Plant Biol 17:558–566
Korpelainen H, Virtanen V (2003) DNA fingerprinting of mosses. J Forensic Sci 48:804–807
Kraaijeveld K, de Weger LA, Ventayol García M, Buermans H, Frank J, Hiemstra PS et al (2015) Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. Mol Ecol Resour 15:8–16
Kress WJ, Erickson DL (2007) A two locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding *trnH-psbA* spacer region. PLoS One 2:e508
Kress JW, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci U S A 102:8369–8374
Lancia M, Conforti F (2013) The use of *Leptodyctium riparium* (Hedw) Warnst. in the estimation of minimum post-mortem interval. J Forensic Sci 58:239–242
Lane MA, Anderson LC, Barkley TM, Bock JH, Gifford EM, Hall DW, Norris DO, Rost TL, Stern WL (1990) Forensic botany – plants, perpetrators, pests, poisons, and pot. Bioscience 40:34–39
Lee HC, Ladd C, Bourke MT et al (1994) DNA typing in forensic science. I. Theory and background. Am J Forensic Med Pathol 15:269–282
Linacre A, Graham D (2002) Role of molecular diagnostics in forensic science. Expert Rev Mol Diagn 2:346–353
Linacre A, Tobe SS (2011) An overview to the investigative approach to species testing in wildlife forensic science. Investig Genet 2:2. https://doi.org/10.1186/2041-2223-2-2
Little DP (2014) A DNA mini-barcode for land plants. Mol Ecol Resour 14(3):437–446. https://doi.org/10.1111/1755-0998.12194
Mathewes RW (2006) Forensic palynology in Canada: an overview with emphasis on archaeology and anthropology. Forensic Sci Int 163:198–203. https://doi.org/10.1016/j.forsciint.2006.06.069
Matsuki Y, Tateno R, Shibata M, Isagi Y (2008) Pollination efficiencies of flower-visiting insects as determined by direct genetic analysis of pollen origin. Am J Bot 95:925–930
Mello ICT, Ribeiro ASD, Dias VHG, Silva R, Sabino BD, Garrido RG et al (2016) A segment of *rbcL* gene as a potential tool for forensic discrimination of *Cannabis sativa* seized at Rio de Janeiro, Brazil. Int J Legal Med 130(2):353–356. https://doi.org/10.1007/s00414-015-1170-x
Mercuri AM (2015) Applied palynology as a trans-disciplinary science: the contribution of aerobiology data to forensic and palaeoenvironmental issues. Aerobiologia 31:323–339
Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. PLoS Biol 3:e422
Mildenhall DC (2006a) An unusual appearance of a common pollen type indicates the scene of the crime. Forensic Sci Int 163:236–240

Mildenhall DC (2006b) Hypericum pollen determines the presence of burglars at the scene of a crime: an example of forensic palynology. Forensic Sci Int 163:231–235
Mildenhall DC, Wiltshire PE, Bryant VM (2006) Forensic palynology: why do it and how it works. Forensic Sci Int 163:163–172
Montali E, Mercuri AM, Trevisan Grandi G, Accorsi CA (2006) Towards a crime pollen calendar—pollen analysis on corpses throughout one year. Forensic Sci Int 163:211–223
Morgan RM, Allen E, King T, Bull PA (2014a) The spatial and temporal distribution of pollen in a room: forensic implications. Sci Justice 54:49–56
Morgan RM, Flynn J, Sena V, Bull PA (2014b) Experimental forensic studies of the preservation of pollen in vehicle fires. Sci Justice 54:141–145
Muellner A, Schaefer H, Lahaye R (2011) Evaluation of candidate DNA barcoding loci for economically important timber species of the mahogany family (Meliaceae). Mol Ecol Resour 11:450–460
Newmaster SG, Fazekas AJ, Ragupathy S (2006) DNA barcoding in the land plants: evaluation of *rbcL* in a multigene tiered approach. Can J Bot 84:335–341
Ogden R, Linacre A (2015) Wildlife forensic science: a review of genetic geographic origin assignment. Forensic Sci Int Genet 18:152–159. https://doi.org/10.1016/j.fsigen.2015.02.008
Petterd CI, Hamshere J, Stewart S, Brinch K, Masi T, Roux C (1999) Glass particles foundin the clothing of members of the public in south-eastern Australia—a survey. Forensic Sci Int 103:193–198
Pimm SL, Joppa LN (2015) How many plant species are there, where are they, and at what rate are they going extinct? Ann Mo Bot Gard 100:170–176
Pompanon F, Deagle BE, Symondson WOC, Brown DS, Jarman SN, Taberlet P (2012) Who is eating what: diet assessment using next generation sequencing. Mol Ecol 21:1931–1950
Pounds CA, Smalldon KW (1975a) The transfer of fibres between clothing materials during simulated contacts and their persistence during wear: Part I—fibre transference. J Forensic Sci Soc 15:17–27
Pounds CA, Smalldon KW (1975b) The transfer of fibres between clothing materials during simulated contacts and their persistence during wear: Part II. J Forensic Sci Soc 15:29–37
Pounds CA, Smalldon KW (1975c) The transfer of fibres between clothing materials during simulated contacts and their persistence during wear: Part III—a preliminary investigation of the mechanisms involved. J Forensic Sci Soc 15:197–207
Quatrehomme G, Lacoste A, Bailet P, Grevin G, Ollier A (1997) Contribution of microscopic plant anatomy to postmortem bone dating. J Forensic Sci 42:140–143
Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System (www.barcodinglife.org). Mol Ecol Notes 7:355–64
Richardson RT, Lin C-H, Sponsler DB, Quijia JO, Goodell K, Johnson RM (2015) Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. Appl Plant Sci 3:1400066. https://doi.org/10.3732/apps.1400066
Sass C, Little D, Stevenson DW, Specht CD (2007) DNA barcoding in the Cycadales: testing the potential of proposed barcoding markers for species identification of Cycads. PLoS One 2: e1154
Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopaedia of life: an introduction to DNA barcoding. Philos Trans R Soc Lond B 360:1805–1811
Shendure J, Ji H (2008) Next-generation DNA sequencing. Nat Biotechnol 26:1135–1145
Shokralla S, Gibson JF, Nikbakht H, Janzen DH, Hallwachs W, Hajibabaei M (2014) Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. Mol Ecol Resour 14(5):892–901
Siver PA, Lord WD, McCarthy DJ (1994) Forensic limnology: the use of freshwater algal community ecology to link suspects to an aquatic crime scene in southern New England. J Forensic Sci 39:847–853
Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. Mol Ecol 21:2045–2050

Taylor B, Skene KR (2003) Forensic palynology: spatial and temporal considerations of spora deposition in forensic investigations. Aust J Forensic Sci 35:193–204
Tsai LC, Yu YC, Hsieh HM, Wang JC, Linacre A, Lee JCI (2006) Species identification using sequences of the trnL intron and the *trnL-trnF* IGS of chloroplast genome among popular plants in Taiwan. Forensic Sci Int 164:193–200
Tsai LC, Yu YC, Hsieh HM, Liu KL, Linacre A, Lee JCI (2008) Bidens identification using the non-coding regions of chloroplast genome and nuclear ribosomal DNA. Forensic Sci Int Genet 2:35–40
Valentini A, Miquel C, Nawaz MA, Bellemain E, Coissac E, Pompanon F et al (2009) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. Mol Ecol Resour 9:51–60
Valentini A, Miquel C, Taberlet P (2010) DNA barcoding for honey biodiversity. Diversity 2:610–617
Virtanen HE, Bjerknes R, Cortes D, Jørgensen N, Rajpert-De Meyts E, Thorsson AV et al (2007) Cryptorchidism: classification, prevalence and long term consequences. Acta Paediatr 96:611–616
Walsh KAJ, Horrocks M (2008) Palynology: its position in the field of forensic science. J Forensic Sci 53:1053–1060
Ward J, Peakall R, Gilmore SR, Robertson J (2005) A molecular identification system for grasses: a novel technology for forensic botany. Forensic Sci Int 152:121–131
Whiltshire E (2009) Forensic ecology, botany and palynology: some aspects of their role in criminal investigation. In: Ritz K, Dawson L, Miller D (eds) Criminal and environmental soil forensics. Springer, Dordrecht, pp 129–149
Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. Syst Biol 54:844–851
Willey P, Heilman A (1987) Estimating time since death using plant roots and stems. J Forensic Sci 32:1264–1270
Wiltshire PEJ (2006) Consideration of some taphonomic variables of relevance to forensic palynological investigation in the United Kingdom. Forensic Sci Int 163:173–182. https://doi.org/10.1016/j.forsciint.2006.07.011
Wolodarsky-Franke A, Lara A (2005) The role of "forensic" dendrochronology in the conservation of alerce (Fitzroya cupressoides ((Molina) Johnston)) forests in Chile. Dendrochronologia 22:235–240
Yaman B, Akkemik U (2009) The use of dendrochronological method in dating of illegal tree cuttings in Turkey: a case study. Balt For 15:122–126
Yoon CK (1993) Botanical witness for the prosecution. Science 260:894–895
Young JM, Weyrich LS, Cooper A (2014) Forensic soil DNA analysis using high-throughput sequencing: a comparison of four molecular markers. Forensic Sci Int Genet 13:176–184
Zavada MS, McGraw SM, Miller MA (2007) The role of clothing fabrics as passive pollen collectors in the north-eastern United States. Grana 46:285–291
Zaya N, Mary VA (2012) Plant genetics for forensic applications. In: Sucher NJ et al (eds) Plant DNA fingerprinting and barcoding: methods and protocols. Methods in molecular biology, vol 862, pp 35–52
Zhou LJ, Pei KQ, Zhou B, Ma KP (2007) A molecular approach to species identification of Chenopodiaceae pollen grains in surface soil. Am J Bot 94:477–481

A Molecular Assessment of Red Algae with Reference to the Utility of DNA Barcoding

Zahid Hameed Siddiqui, Zahid Khorshid Abbas, Khalid Rehman Hakeem, Mather Ali Khan, and Abdul Ilah

Abstract The ecological and commercial importance of red algae is of high value, the estimated cost seaweed industry produces is US$10 billion. The species which are exploited most are the members of Rhodophyta (*Eucheuma/Kappaphycus*, *Porphyra*, and *Gracilaria*). In order to understand the distribution of seaweed, their identification is necessary which is generally based on morphological characteristics, often resulting in wrong identification of species. DNA barcoding can be used as a contemporary tool for species identification. It can resolve many intrinsic problems of morphological taxonomy, only a small amount of tissue is required for species identification, and the samples can be examined at all stages of development. The application of DNA barcoding can be used in identification of invasive and endangered species along with conservation biology. In case of red alga, DNA barcoding proved to be beneficial for the recognition of high-yielding agar strain as well as for cryptic species identification. In this study, several identification-based problems of red algae have been discussed by using different intraspecific markers such as cox1, cox3, and cox2-3 spacer and rbcL and rbcL-rbcS spacer. As per the available data, the mitochondrial markers, gene cox1 is more effective than rbcL for the measurement of red algal genetic diversity.

Keywords DNA barcoding · Red algae · Cox1 · Cox2-3 · rbcL

Z. H. Siddiqui (✉) · Z. K. Abbas
Department of Biology, Faculty of Science, University of Tabuk, Tabuk, Kingdom of Saudi Arabia

K. R. Hakeem
Department of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia

M. A. Khan
Bond Life Sciences Center, University of Missouri-Columbia, Columbia, MO, USA

A. Ilah
Faculty of Medical Technology, Omar Al Mukhtar University, Al Bayda, Libya
e-mail: mdabdul.ilah@omu.edu.ly

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_10

Abbreviations

AFLP	Amplified fragment length polymorphism
BOLD	Barcode of Life Database
CBOL	Consortium for the Barcode of Life
cox1	Mitochondrial cytochrome oxidase subunit 1
cox2	Mitochondrial cytochrome oxidase subunit 2
cox3	Mitochondrial cytochrome oxidase subunit 3
DNA	Deoxyribonucleic acid
ISSR	Inter-simple sequence repeats
ITS-1 and ITS-2	The internal transcribed spacers
RAPD	Random amplified polymorphic DNA
rbcL	Large subunit of the ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo)
rbcL-rbcS	The plastid-encoded RuBisCO spacer
rDNA (SSU rDNA)	The nuclear small subunit
RFLP	Restriction fragment length polymorphism

1 Introduction

In this era of globalization, human activities are causing climate change, pollution, coastal degradation, and introduction of alien and invasive species. During these changes, seaweed populations are also affected severely irrespective of nation and continents. Among the seaweed, the members of Rhodophyta are very important, ecologically as well as commercially. Therefore, in order to understand the distribution of seaweed, identification of organisms is necessary which is generally based on morphological characteristics such as size, shape, and color of the organism's parts. This is the traditional way of identification which is little complex because of its three major limitations (Hebert et al. 2003; Pires and Marinoni 2010). Firstly, the key on the basis of morphological characteristics is not complete or not available for all taxa due to lack of proper taxonomic description (May 1988) as some groups (flowering plants and vertebrates) are better studied than others (algae, nematodes). Secondly, the morphological characteristics and their complexity depend on the group under consideration; therefore, sometimes the identification keys need well-trained taxonomists, who are becoming very rare and are not always available for regular identifications: this is referred to as taxonomic impediment. Thirdly, the organism to be identified may be too small or at a developmental stage, where even trained taxonomists face difficulty in characterization and identification. In addition to these, the morphology-based methods are time consuming and do not provide classification up to the species level (Rindi et al. 2008; Packer et al. 2009). With the advent of molecular biology, it has been revealed that a large number of biological species show genetic divergence without accompanying morphological disparities and therefore cannot be identified by traditional methods. The

recognition of these cryptic species is a major challenge to modern taxonomy (Heinrichs et al. 2011). During the last two decades, small segments of DNA, called as DNA barcodes, have been developed for differentiation of species (Yow et al. 2013). In short, DNA barcoding is based on small, single marker sequence similarity comparison to classify species (Hebert et al. 2003). The author's further debated that by incorporating DNA barcoding, the unknown biodiversity can be classified quickly and reliably than traditional taxonomic methods alone (Hebert et al. 2004). The identification of species using DNA barcoding has been effectively documented in animals such as spiders (Barrett and Hebert 2005), fish (Ward et al. 2005), birds (Hebert et al. 2004), and rodents (Robins et al. 2007) as well as in plants (Chase et al. 2005; Fazekas et al. 2012), fungi (Seena et al. 2010), and algae (Saunders 2008).

In animals, with about 600 base pairs, the 5′end portion of the mitochondrial gene cytochrome oxidase 1 (cox1) was selected as the main molecular marker (Hebert et al. 2003). Later on, the same gene was adapted in Rhodophyceae (Saunders 2005) and Phaeophyceae (McDevit and Saunders 2010) for molecular classification. Using cox1, a large number of articles have been published for red as well as brown algae, which indicates the usefulness of this marker for species-level identification, unraveling cryptic diversity and variations within species (Brodie et al. 2008). However, the use of universal plastid amplicon (domain V of the plastid large subunit rDNA) was proposed by Sherwood and Presting (2007) as a DNA barcode in photosynthetic organisms for rapid species identification. In this chapter, we will discuss the ecological and economic importance of red algae and DNA barcoding and its use in red algae as a case study.

2 Ecological and Economic Importance of Red Algae

The Rhodophyta is a separate taxon characterized by the accessory photosynthetic pigments phycocyanin, phycoerythrin, and allophycocyanins arranged in phycobilisomes (Woelkerling 1990). It is commonly known as red algae and encompasses a large number of species with different types of body shape, ranging from unicellular-filamentous (simple) to the blade-pseudoparenchymatous (most complex) (Cole and Sheath 1995). As per Guiry (2012), in AlgaeBase there are around 6131 species of red algae, the total number of estimated species which are described is 7000, and the species which are undescribed are 7000. This information tells us the need of identification and classification of this important taxon (Rhodophyta) which is very important from environmental as well as from economic perspective. There are several genera of red algae which are found in freshwaters, but the majority of red algae are of marine nature; they are generally reported on tropical and temperate marine shores. At these shores, they play the role of "keystone species" by building and maintaining coral reefs (Freshwater 2000). This gives banded coral, giant clams, clownfish, shrimps, and other animals a reliable habitat where they can dwell and maintain ecological balance. Further, these algae form flat sheets that fuse and stabilize reef crest and protect reefs from wave damage. There

are fossil evidences which indicate that these coralline red algae play this important role from hundreds of millions of years (Freshwater 2000; Graham et al. 2009).

Recently Rebours et al. (2014) reported the estimated value of seaweed industry, producing US$10 billion. The species which are exploited most are the members of Rhodophyta, (*Eucheuma/Kappaphycus*, *Porphyra*, and *Gracilaria*). Among these species *Kappaphycus* alone produces US$1.3 billion, while the nori market creates US$1.5 billion. *Porphyra yezoensis*, *Gracilaria*, and some other species are grown in mariculture for use as human food; in fact, most of the algae contain amino acids; proteins; carbohydrates; vitamins B1, B2, B12, and C; and carotenoid that are essential for the normal functioning of the human body. It also contains minerals such as potassium, magnesium, iron, selenium, and large amount of iodine, and their fat content is very low (around 0.2–2%), which makes it a good source of balanced nutrition (Lee 2008). *Kappaphycus* along with other genera are cultivated for extraction of gel-forming agar, agarose, and carrageenan. These gelling polysaccharides are commonly used in laboratory cell culture media preparation (Yeong et al. 2008), genomics-/proteomics-based research (Siow et al. 2012), and food processing (Van de Velde et al. 2002). Red algae are also looking for as a source of compounds that can be used against microbial or herbivore attack as well as a potential source for human medicine (Bixler 1996; Villanueva et al. 2008; Graham et al. 2009; Yow et al. 2013). Table 1 lists the medicinal role and industrial uses of some of the important genera of red algae.

During the last 50 years, the algal industry has been increased multifold with value of over US$6 billion (FAO 2014; Loureiro et al. 2015). Further, the cultivation of algae holds much greater potential ranging from integrated fish farming to biofuel production. Therefore, identification of correct species for differential use is the need of the hour, and DNA barcoding seems to be a promising solution for the same.

3 DNA Barcoding

The existence of huge biodiversity in nature is because of the genetic differences among organisms which cannot be recognized by using traditional morphological-based studies; the identification of such hidden species is a huge challenge to modern taxonomy (Heinrichs et al. 2011). DNA barcoding can resolve many intrinsic problems of morphological taxonomy; we are well aware that the numbers of taxonomists are decreasing day by day, and on the other hand, numbers of species are increasing; therefore, in order to manage this situation, molecular biology helps to solve the taxonomical problems. In molecular biology, only a small amount of tissue is required for species identification, and the samples can be examined at all stages of development (Floyd et al. 2002; Savolainen et al. 2005). Nowadays, barcoding is routinely used for multicellular organisms, such as butterflies (Burns et al. 2008), birds (Hebert et al. 2004), and aquatic hyphomycetes (Seena et al. 2010).

Arnot et al. (1993) first used "DNA barcode" while working on *Plasmodium falciparum*, but the use of molecular biology for organism identification was older

Table 1 Medicinal role and industrial uses of some of the genera of Rhodophyta

Medicinal role		
Genus	Uses	
Porphyra	Can be used to treat goiter, bronchitis, tonsillitis, and cough	
Gelidium	Laxative; can be used to treat tracheitis, gastric diseases, and hemorrhoids; can be used to extract agar	
Pterocladiella	Laxative; can be used to treat tracheitis, gastric disease, and hemorrhoids; can be used as an agar extract; can be used to make coating for pills	
Eucheuma	Can be used to treat goiter, tonsillitis, bronchitis, asthma, cough, gastric diseases, and hemorrhoids	
Corallina	Marl, parasiticide	
Gracilaria	Can be used to treat goiter, edema, urinary diseases; can prevent ulcer; can be used to as an agar extract and make coating for pills	
Hypnea	Can be used to treat bronchitis, gastric diseases, and hemorrhoids; can be used to make carragenate	
Chondrus	Can be used to treat bronchitis, tonsillitis, gastric diseases, asthma, and cough; is an adhesive; can be used to make carragenate	
Centroceras	Can be used to treat gastrointestinal intolerance	
Chondria	Ascaricide	
Grateloupia	Ascaricide; lowers blood pressure	
Gloiopeltis	Can be used to treat goiter, tonsillitis, and bronchitis; prevents high blood pressure and scurvy	
Industrial use of red algae		
Phycolloids	Genus	Use
Agar	*Gelidium* *Pterocladiella* *Gracilaria*	Culture medium for microorganism, food industry (jellies, jams, starch substitutes, stabilizer for canned food, gelling agent), textile industry, papermaking, clarifier for wine making, aids in the making of ultrathin separating film, analytical research (electrophoresis), medicine (gastrointestinal intolerance, capsules)
Carrageenan	*Chondrus* *Gigartina* *Eucheuma* *Hypnea*	Food industry (jellies, jams, salad dressing, stabilization agent for canned food, gelling agent), cosmetic industry, medical industry (granulation of pills, base, capsule, gastrointestinal intolerance), paint industry
Furcellaran	*Furcellaria*	Dairy industry
Funoran	*Gloiopeltis*	Papermaking industry, textile industry
Porphyran	*Porphyra*	Adhesive

Compendium of Materia Medica: Source: http://formosa.ntm.gov.tw/seaweeds/english/e/e3_01.asp Accessed on 30 Jan 2017
Li (2003)

(McAndrew and Majumdar 1983; Anderson et al. 1985). Allozymes were the first molecules used for species differentiation by Hubby and Lewontin (1966). However, the recognition of species by DNA barcodes depends upon the target species which should have reasonable genetic differentiation that can be harnessed for separation of species, even where morphological similarities exist. The novelty behind DNA barcode is to find a single segment of DNA which can be used for the identification

of all living taxa. Researchers are still trying to find a single segment of DNA which is suitable for the identification of all taxa; in spite of more than two decades of work in this particular area, single DNA marker is yet to be reported (Peěnikar and Buzan 2014). Based on the amount of work in DNA barcoding, Peěnikar and Buzan (2014) laid down certain qualities of DNA barcodes as stated below.

1. The fragment of DNA must be nearly identical in specimens of the same species but different between individuals of different species.
2. The section must be standardized and the same section should be used in different taxonomic groups.
3. The marker must be robust, with conservative primer binding sites that allow it to be readily amplified and sequenced.

3.1 Practical Approaches of DNA Barcoding

The initial idea behind development of DNA barcodes was for species identification. It is a research tool for taxonomists and helps in species recognition at various stages of growth and development, e.g., in plants from seeds, seedling to plant, whereas in animals from the formation of eggs, larvae, to mature individual (fertile and sterile). It also helps in species identification in case of damaged samples, unisexual species, gut contents, and fecal matter (Kress and Erickson 2012). Further, DNA barcoding has the potential to evaluate the consistency of species with a universal measure of genetic variability based on the barcode sequence data. The application base use of DNA barcoding can be seen in identification of invasive and endangered species along with conservation biology. It can also be used for checking the identity and purity of materials such as biological products, herbal medicines, dietary supplements, and seafood. DNA barcoding also helps to flag species which are new, especially the cryptic species (Hebert et al. 2004). Nowadays, DNA barcodes also used to solve basic evolutionary and ecological problems (Kress et al. 2010) and for the recognition of medically important pathogens and their invertebrate vectors. It can also be used for the identification of species which is used for manufacturing of drugs of natural origin (Peěnikar and Buzan 2014). Besides that, DNA barcoding can be done on the herbarium specimens and museum samples, in order to reconfirm the efficacy of this new science. All over the globe, large numbers of natural history museums and herbaria have huge collections of specimens which are classified on the basis of morphology by experienced taxonomists. The application of DNA barcoding in these samples helps in proper understanding of earlier classification, and simultaneously specimens will be backed up by the DNA barcodes (Ellis 2008; Puillandre et al. 2012).

This method is giving a quick and reliable identity to unrecognized organisms with the use of verified reference material (Erickson and Kress 2012). For this, Consortium for the Barcode of Life (CBOL) has been developed by the efforts of global DNA barcoding. In January 2017, the Barcode of Life Database (BOLD)

contained more than 6.9 million specimen records, with 5.29 million having barcodes belonging to over 262,108 species (BOLD Systems 2017a). Among plant database of BOLD, the data for red alga specimen is highest (47,976 specimen records) as per January 2017. Of these, the number of specimens with sequences is 35,231, and the numbers of specimens with barcodes are 24,372, whereas 3062 species have barcodes (BOLD Systems 2017b); this data is self-explanatory about the importance of red algae.

4 Markers for DNA Barcoding in Red Algae

The genetic diversity of seaweed compels the researchers for its measurement; with the advent of molecular tools and advancement in molecular biology, this work is growing at a very fast rate, and a lot of research articles are published on this subject (Conklin et al. 2009). Genetic markers give data which can be used for the analysis of biogeographic studies, population structure, and their phylogenetic relationships, parentage, and relatedness in the gene flow (Féral 2002). A large number of molecular tools are easily available and are widely used to explain genetic inconsistency of marine macroalgae, such as AFLP (Donaldson et al. 2000; Pang et al. 2010), DNA markers used in allozyme differentiation, RFLP (Kamikawa et al. 2007), ISSR, RAPD (Zhao et al. 2008), and microsatellites (Zhang et al. 2009). Apart from them, a large group of intraspecific markers are also available for measuring genetic diversity of seaweeds.

1. The large subunit of the ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo) plastid gene rbcL (Nam et al. 2000; Gurgel and Fredericq 2004)
2. The plastid-encoded RuBisCO spacer, the rbcL-rbcS (Byrne et al. 2002; Zuccarello et al. 2006)
3. Mitochondrial cytochrome oxidase subunit 3, cox3 (Steel et al. 2000; Coyer et al. 2004; Uwai et al. 2006)
4. Mitochondrial-encoded cox2-3 spacer (Chiasson et al. 2007; Vidal 2008; Vis et al. 2010)
5. Mitochondrial cytochrome oxidase subunit 1 gene, cox1 (Sherwood 2008; Kim et al. 2010; Yow et al. 2011)
6. The nuclear small subunit rDNA (SSU rDNA) and the internal transcribed spacers, ITS-1 and ITS-2 (Cho et al. 2007; Lindstrom 2008; Moniz and Kaczmarska 2010)

For measuring the genetic diversity of red alga, the usefulness of mitochondrial markers, the cox1 (Yang et al. 2007; Yow et al. 2011) and the cox2-3 spacer (Zuccarello and West 2002), has been reported. As per Saunders (2005), the DNA region of mitochondria gives positive results for species identification of red macroalgae, whereas Geraldino et al. (2006) and Kim et al. (2010) reported gene cox1 as an ideal marker for DNA barcoding of red algae. The gene cox1 also helps in elucidating of the cryptic diversity of red algae (Robba et al. 2006), whereas the

cox2-3 spacer was proven beneficial for the phytogeographical study of rhodophytes (Andreakis et al. 2007; Vis et al. 2008).

5 DNA Barcoding in Red Algae

Due to its ecological and economic importance, lot of DNA barcoding work has been done and undergoing in red algae. As stated above, *Porphyra* are commercially edible seaweeds available in different countries (Turner 2003; Nelson and Broom 2005; Ruangchuay and Notoya 2007); therefore, its accurate identification is very necessary before the establishment of commercial cultivation. Because there are only few morphological characteristics that can be used objectively to identify species, this group of algae has defied taxonomists for a long time. To aggravate the problem, individuals of the same species can be different morphologically (phenotypic plasticity), and individuals of different species can be morphologically similar but highly divergent genetically (Neefus et al. 2002; Milstein and Oliveira 2005; Brodie et al. 2007; Lindstrom 2008).

In Brazil, Milstein et al. (2012) performed an elaborated study on the macroalgal flora of São Paulo state in a biodiversity project. The authors sequenced cox1 and cox2-3 spacer and UPA as three DNA barcode markers for *Porphyra* species from Brazil. The three markers help in the recovery of a cryptic species (*Porphyra* sp. 77) and also clustered two different species (*P. drewiana* and *P. spiralis*) that were synonymized. Further, they reported that varieties such as *P. acanthophora* and *P. spiralis* are only differing morphologically and there is no sequence divergence in three studied molecular markers. In *Gelidium elegans*, the plastid gene rbcL and nuclear 18S ITS genes have been analyzed in samples from Japan and Taiwan (Freshwater et al. 1995; Shimada et al. 2000). In another report, sequences of rbcL, photosystem I P700 chlorophyll a apoprotein A (psaA), and cox1 gene (Kim et al. 2011) have shown that *G. elegans* demonstrates high genetic diversity as compared to other red algal species. Recently, Kim et al. (2012) assess the genetic structure of the Korean species *G. elegans* by analyzing cox1 gene from 272 individuals collected from 36 locations. The authors reported 34 haplotypes, which were unique; the nucleotide and haplotype diversities of cox1 within *G. elegans* were 0.711 ± 0.028 (H) and 0.00736 ± 0.00038 (π), respectively.

In BOLD Systems (2017c), among family Rhodomelaceae, the specimen record of *Polysiphonia* is the highest, i.e., 1456, whereas for the same genus, specimens with sequences and barcodes recorded are 780 and 553, respectively. The phenotypic plasticity in *Polysiphonia* species is very high (Kim et al. 2000); therefore, species identification is very difficult or nearly impossible (Kim and Yang 2006). On the basis of rbcL sequences, *Polysiphonia morrowii* has been reported as an alien species in the Southern Pacific Ocean (Chile and New Zealand) (Kim et al. 2004; Mamoozade and Freshwater 2011). As per the previous records, this species naturally belongs to Northwest Pacific Ocean and has been recorded in China (Segi 1951), Far East Russia (Perestenko 1980), Japan (Kudo and Masuda 1992), and South Korea (Kim

et al. 1994). Earlier, the presence of *P. morrowii* was debatable on the basis of its morphology and anatomy in New Zealand (Nelson and Maggs 1996), in the Mediterranean Sea (Erdŭgan et al. 2009), and putatively in the North Sea as *P. senticulosa* Harvey (Maggs and Stegenga 1999). Geoffroy et al. 2012 sampled 110 specimens of *Polysiphonia* and used a DNA barcoding approach for their identification at the species level. A total of 110 rbcL sequences were generated, and the alignment covered 1225 bp with 430 variable sites. The sequences generated in the present study were resolved in the lineage containing the *P. morrowii* sequences from GenBank. Among the 105 individuals of *P. morrowii* collected along the coast of Brittany, three haplotypes were found suggesting several introduction events. *P. stricta* joined *P. pacifica* and were resolved as a sister group of *P. morrowii*. Hence, the authors demonstrated the presence of *P. morrowii* as an alien species in the North-Eastern Atlantic with the help of DNA barcoding.

Gracilaria changii is a very important commercial species in Peninsular Malaysia with high genetic diversity; for better agar yield and high growth rate, it is required to identify the best strain, as these features are species or strain specific. Beside that for better conservation and management of various species, strain identification is very necessary. Information on intraspecific genetic diversity of this economically important species is scarce (Yow et al. 2011). Keeping these points Yow et al. (2013) use mitochondrial cox1 gene and cox2-3 spacer for measuring genetic diversity of *G. changii*. The authors reported that cox1 gene is a potential marker for defining intraspecific genetic variation in red algae and cox1 marker is more variable as compared to cox2-3 spacer. In another report, Robba et al. (2006) performed a comparative analysis between plastid RuBisCO spacer and cox1 from numerous samples of red algae and concluded that cox1 is a more sensitive marker for revealing population structure and the hidden diversity of red algal species. Yang et al. (2008) reported similar results in *G. vermiculophylla* and putative relatives their results suggested that cox1 is a valuable molecular marker within species and for haplotype analyses. Recently Yang and Kim (2015) performed molecular analyses of Gracilariaceae (Rhodophyta) for identification from the Asia-Pacific region. The authors analyzed mitochondrial cox1 and plastid rbcL genes as marker; the results revealed 22 species as distinct entity including five unknown and two new distribution records. On comparative analysis of cox1 and rbcL sequence data, cox1 showed more variation than rbcL data, allowing the perfect discrimination of species.

The Philippines is the leading producer and holds almost 70% of the world's semi-refined carrageenan supply by culturing *Eucheuma denticulatum* (N. L. Burman) F.S. Collins and Hervey and *Kappaphycus sp.*, (Llana 1991; Villanueva et al. 2008). The farming of these two species outside the Philippines has been profitable in only a few countries (Hurtado and Agbayani 2002; McHugh 2003). Due to shortage of distinguished morphological characteristics for recognition and high morphological plasticity within the genera, *Eucheuma* and *Kappaphycus* have created misperception about the distributions and spread of three introduced eucheumoid species in Hawaii (Conklin et al. 2009). The authors employed DNA barcoding and used three molecular markers, namely, partial nuclear

28S rRNA, partial plastid 23S rRNA, and mitochondrial 5′ cox1, in order to identify *Eucheuma* and *Kappaphycus* samples from Hawaii.

Another red algae *Grateloupia* of family Halymeniaceae has wide geographical distribution, it ranges from tropical to warm temperate regions of the world. There is a very high morphological similarity among various species of *Grateloupia,* especially between *G. elliptica* and *G. lanceolata*. These two species possess blade-like thalli with leather texture and cruciately divided tetrasporangia; as a result of similar appearance, there is a recurrent confusion which leads to difficulty in the identification. Yang et al. (2013) analyzed the relationships between two species by using plastid rbcL and mitochondrial cox1 gene as a biomarker. The rbcL sequence analyses backed the difference between two species, with interspecific divergence of 3.7–4.6%. The genetic diversity of cox1 gene within species was estimated to be 0–0.3% in *G. elliptica* and 0–1.0% in *G. lanceolata*, respectively. Finally, the authors reported that cox1 gene is more effective than rbcL. Wolf et al. (2011) reported the presence of exotic *Hypnea flexicaulis* in the Mediterranean Sea by using DNA barcoding employing rbcL and cox1 gene along with morphological analysis.

6 Conclusion

It is a well-known fact that genera of red algae are very important commercially as well as ecologically. Due to human activities, there is introduction of new species which are exotic and invasive in nature, for ecological conservation species identification is necessary because of symbiotic nature of these algae which are species specific. Further, for better agar yield and growth, species identification is mandatory. We also know that because of limited characteristics and phenotypic plasticity, identification on the basis of morphology is doubtful by this age-old practice. This chapter deals with resolving taxonomical problem of red algae by means of DNA barcoding; it is a reliable tool when species identification is required in a quick manner. Some cases of red algae are described here; on that basis, it is concluded that cox1 is a more sensitive marker for revealing population structure and the hidden diversity.

Conflict of Interest The authors declare that there is no conflict of interest.

References

Anderson NL, Parish NM, Richardson JP, Pearson TW (1985) Comparison of African trypanosomes of different antigenic phenotypes, subspecies and life cycle stages by two-dimensional gel electrophoresis. Mol Biochem Parasitol 16(3):299–314

Andreakis N, Procaccini G, Maggs C, Kooidtra WHCF (2007) Phylogeography of the invasive seaweed *Asparagopsis* (Bonnemaisoniales, Rhodophyta) reveals cryptic diversity. Mol Ecol 16:2285–2299

Arnot DE, Roper C, Bayoumi RA (1993) Digital codes from hypervariable tandemly repeated DNA sequences in the *Plasmodium falciparum* circumsporozoite gene can genetically barcode isolates. Mol Biochem Parasitol 61(1):15–24

Barrett RD, Hebert PD (2005) Identifying spiders through DNA barcodes. Can J Zool 83 (3):481–491

Bixler HJ (1996) Recent developments in manufacturing and marketing carrageenan. Hydrobiologia 326/327:35–57

BOLD Systems (2017a) http://www.boldsystems.org/index.php/TaxBrowser_Home

BOLD Systems (2017b) http://www.boldsystems.org/index.php/Taxbrowser_Taxonpage?taxid=48327

BOLD Systems (2017c) http://www.boldsystems.org/index.php/Taxbrowser_Taxonpage?taxid=55735

Brodie J, Bartsch I, Neefus C, Orfinidis S, Bray T, Mathieson A (2007) New insights into the cryptic diversity of the North Atlantic–Mediterranean '*Porphyra leucosticta*' complex: *P. olivii* sp. nov. and *P. rosengurttii* (Bangiales, Rhodophyta). Eur J Phycol 42:3–28

Brodie J, Mortensen AM, Ramirez ME, Russell S, Rinkel B (2008) Making the links: towards a global taxonomy for the red algal genus *Porphyra* (Bangiales, Rhodophyta). J Appl Phycol 20:939–949

Burns JM, Janzen DH, Hajibabaei M, Hallwachs W, Hebert PDN (2008) DNA barcodes and cryptic species of skipper butterflies in the genus Perichares in Area de Conservacion Guanacaste, Costa Rica. Proc Natl Acad Sci USA 105:6350–6355

Byrne K, Zuccarello GC, West J, Liao ML, Kraft GT (2002) *Gracilaria species* (Gracilariaceae, Rhodophyta) from southeastern Australia, including a new species, *Gracilaria perplexa* sp. nov.: morphology, molecular relationships and agar content. Phycol Res 50:295–312

Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP, Haidar N, Savolainen V (2005) Land plants and DNA barcodes: short-term and long-term goals. Philos Trans R Soc Lond Ser B Biol Sci 360(1462):1889–1895

Chiasson WB, Johanson KG, Sherwood AR, Vis ML (2007) Phylogenetic affinities of the form taxon *Chantransia pygmaea* (Rhodophyta) specimens from the Hawaiian Islands. Phycologia 46:257–262

Cho GY, Kogame K, Kawai H, Boo SM (2007) Genetic diversity of *Scytosiphon lomentaria* (Scytosiphonaceae, Phaeophyceae) from the Pacific and Europe based on RuBisCo large subunit and spacer, and ITS nrDNA sequences. Phycologia 46:657–665

Cole KM, Sheath RG (1995) Biology of the Red Sea weeds. Cambridge University Press, New York

Compendium of Materia Medica. Source: http://formosa.ntm.gov.tw/seaweeds/english/e/e3_01.asp. Accessed 30 Jan 2017

Conklin KY, Kurihara AA, Sherwood AR (2009) A molecular method for identification of the morphologically plastic invasive algal genera *Eucheuma* and *Kappaphycus* (Rhodophyta, Gigartinales) in Hawaii. J Appl Phycol 21:691–699

Coyer JA, Hoarau G, Stam WT, Olsen JL (2004) Geographically specific heteroplasmy of mitochondrial DNA in the seaweed, *Fucus serratus* (Heterokontophyta, Phaeophyceae, Fucales). Mol Ecol 13:1323–1326

Donaldson SL, Chopin T, Saunders GW (2000) An assessment of the AFLP method for investigating populations structure in the red alga *Chondus crispus*. J Appl Phycol 12:25–35

Ellis R (2008) Rethinking the value of biological specimens: laboratories, museums and the Barcoding of Life Initiative. Mus Soc 6(2):172–191

Erdŭgan H, Akı C, Acar O, Dural B, Aysel V (2009) New record for the east Mediterranean, Dardanelles (Turkey) and its distribution: *Polysiphonia morrowii* Harvey (Ceramiales, Rhodophyta). Turk J Fish Aquat Sci 9:231–232

Erickson DL, Kress WJ (2012) Future directions. In: Kress JW, Erickson DL (eds) DNA barcodes: methods and protocols, Methods in molecular biology, vol 858. Humana Press, New York, pp 459–465

Fazekas AJ, Kuzmina ML, Newmaster SG, Hollingsworth PM (2012) DNA barcodingmethods for land plants. In: Kress JW, Erickson DL (eds) DNA barcodes: methods and protocols, Methods in molecular biology, vol 858. Springer, New York, pp 223–252

Féral JP (2002) How useful are the genetic markers in attempts to understand and manage marine biodiversity? J Exp Mar Biol Ecol 268:121–145

Floyd R, Abebe E, Papert A, Blaxter M (2002) Molecular barcodes for soil nematode identification. Mol Ecol 11:839–850

Food and Agriculture Organization of the United Nations (2014) Fisheries and aquaculture information and statistics services. http://www.fao.org/figis/. Accessed 2 Feb 2017

Freshwater DW (2000) Rhodophyta. Red Algae. Version 24 March 2000 (under construction). http://tolweb.org/Rhodophyta/2381/2000.03.24 *in* The Tree of Life Web Project

Freshwater W, Fredericq S, Hommersand MH (1995) A molecular phylogeny of the *Gelidiales* (Rhodophyta) based on analysis of plastid rbcL nucleotide sequences. J Phycol 31:616–632

Geoffroy A, Gallc LL, Destombe C (2012) Cryptic introduction of the red alga *Polysiphonia morrowii* Harvey (Rhodomelaceae, Rhodophyta) in the North Atlantic Ocean highlighted by a DNA barcoding approach. Aquat Bot 100:67–71

Geraldino PJL, Yang EC, Boo SM (2006) Morphology and molecular phylogeny of *Hypnea flexicaulis* (Gigartinales, Rhodophyta) from Korea. Algae 21:417–423

Graham LE, Graham JM, Wilcox JW (2009) Red algae. In: Algae, 2nd edn. Pearson Benjamin Cummings, San Francisco, CA

Guiry MD (2012) How many species of algae are there? J Phycol 48:1057–1063

Gurgel CFD, Fredericq S (2004) Systematics of Gracilariaceae (Gracilariales, Rhodophyata): a critical assessment based on rbcL sequence analyses. J Phycol 40:138–159

Hebert PDN, Cywinska A, Ball SL, de Waard JR (2003) Biological identifications through DNA barcodes. Proc Biol Sci 270:313–321

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. PLoS Biol 2(10):e312

Heinrichs J, Kreier HP, Feldberg K, Schmidt AR, Zhu RL, Shaw B, Shaw AJ, Wissemann V (2011) Formalizing morphologically cryptic biological entities: new insights from DNA taxonomy, hybridization, and biogeography in the leafy liverwort *Porella platyphylla* (Jungermanniopsida, Porellales). Am J Bot 98(8):1252–1262

Hubby JL, Lewontin RC (1966) A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. Genetics 54:577–594

Hurtado AQ, Agbayani RF (2002) Deep-sea farming of *Kappaphycus* using the multiple raft, long-line method. Bot Mar 45:438–444

Kamikawa R, Masuda I, Oyawa K, Yoshimatsu S, Sako Y (2007) Genetic variation in mitochondrial genes and intergenic spacer region in harmful algae *Chattonella species*. Fish Sci 73:871–880

Kim MS, Yang EC (2006) Taxonomy and phylogeny of *Neosiphonia japonica* (Rhodomelaceae, Rhodophyta) based on rbcL and cpeA/B gene sequences. Algae 2:287–294

Kim MS, Lee IK, Boo SM (1994) Morphological studies of the red alga *Polysiphonia morrowii* Harvey on the Korean coast. Korean J Phycol 9:185–192

Kim MS, Maggs C, McIvor L, Guiry M (2000) Reappraisal of the type species of *Polysiphonia* (Rhodomelaceae, Rhodophyta). Eur J Phycol 35:83–92

Kim MS, Yang EC, Mansilla A, Boo SM (2004) Recent introduction of *Polysiphonia morrowii* (Ceramiales, Rhodophyta) to Punta Arenas, Chile. Bot Mar 47:389–394

Kim SY, Weinberger F, Boo SM (2010) Genetic diversity hints at a common donor region of the invasive Atlantic and Pacific populations of *Gracilaria vermiculophylla* (Rhodophyta). J Phycol 46:1346–1349

Kim KM, Hwang IK, Park JK, Boo SM (2011) A new agarophyte species, *Gelidium eucorneum* sp. nov. (Gelidiales, Rhodophyta), based on molecular and morphological data. J Phycol 47:904–910

Kim KM, Hoarau GG, Boo SM (2012) Genetic structure and distribution of *Gelidium elegans* (Gelidiales, Rhodophyta) in Korea based on mitochondrial cox1 sequence data. Aquat Bot 98:27–33

Kress WJ, Erickson DL (2012) DNA barcodes: methods and protocols. Methods Mol Biol 858:3–8

Kress WJ, Erickson DL, Swenson NG, Thompson J, Uriarte M, Zimmerman JK (2010) Advances in the use of DNA barcodes to build a community phylogeny for tropical trees in a Puerto Rican forest dynamics plot. PLoS One 5(11):e15409

Kudo T, Masuda M (1992) Taxonomic features of *Polysiphonia morrowii* Harvey (Ceramiales, Rhodophyta). Korean J Phycol 7:13–26

Lee RE (2008) Phycology, vol 107. Cambridge University Press, Cambridge

Li S (2003) In Xiwen L (ed) Compendium of Materia Medica: Bencao Gangmu. Foreign Languages Press, Beijing

Lindstrom SC (2008) Cryptic diversity, biogeography and genetic variation in Northeast Pacific species of *Porphyra sensu lato* (Bangiales, Rhodophyta). J Appl Phycol 20:501–512

Llana EG (1991) Production and utilisation of seaweeds in the Philippines. INFOFISH Int 1/91:12–17

Loureiro R, Gachon CMM, Rebours C (2015) Seaweed cultivation: potential and challenges of crop domestication at an unprecedented pace. New Phytol 206(2):489–492

Maggs CA, Stegenga H (1999) Red algal exotics on North Sea coasts. Helgoländer Meeresun 52:243–258

Mamoozade NR, Freshwater DW (2011) Taxonomic notes on Caribbean *Neosiphonia* and *Polysiphonia* (Ceramiales, Florideophyceae): five species from Florida, USA and Mexico. Bot Mar 54:269–292

May RM (1988) How many species are there on earth? Science 241:1441–1449

McAndrew BJ, Majumdar KC (1983) Tilapia stock identification using electrophoretic markers. Aquaculture 30(1–4):249–261

McDevit DC, Saunders GW (2010) A DNA barcode examination of the *Laminariaceae* (Phaeophyceae) in Canada reveals novel biogeographical and evolutionary insights. Phycologia 49:235–248

McHugh DJ (2003) A guide to the seaweed industry. Technical Report 441. Food and Agriculture Organization of the United Nations, Rome, Italy

Milstein D, Oliveira MC (2005) Molecular phylogeny of Bangiales (Rhodophyta) based on small subunit rDNA sequencing: emphasis on Brazilian *Porphyra species*. Phycologia 44:212–221

Milstein D, Medeiros AS, Oliveira EC, Oliveira MC (2012) Will a DNA barcoding approach be useful to identify *Porphyra species* (Bangiales, Rhodophyta)? A case study with Brazilian taxa. J Appl Phycol 24:837–845

Moniz MBJ, Kaczmarska I (2010) Barcoding of diatoms: nuclear encoded ITS revisited. Protist 161:7–34

Nam KW, Maggs CA, McIvor L, Stanhope MJ (2000) Taxonomy and phylogeny of Osmunder (Rhodomelaceae, Rhodophyta) in Atlantic. Eur J Phycol 36:759–772

Neefus CD, Mathieson AC, Klein AS, Teasdale B, Gray T, Yarish C (2002) *Porphyra birdiae* sp. nov. (Bangiales, Rhodophyta): a new species from the northwest Atlantic. Algae 17:203–216

Nelson WA, Broom JES (2005) Contributions of molecular biology to understanding systematics and phylogeny in the order Bangiales. Nat Hist Res Spec Issue 8:1–12

Nelson WA, Maggs CA (1996) Records of adventive marine algae in New Zealand: *Antithamnionella ternifolia*, *Polysiphonia senticulosa* (Ceramiales, Rhodophyta), and *Striaria attenuata* (Dictyosiphonales, Phaeophyta). NZ J Mar Freshw Res 30:449–453

Packer L, Gibbs J, Sheffield C, Hanner R (2009) DNA barcoding and the mediocrity of morphology. Mol Ecol Resour 9(Suppl S1):42–50

Pang QQ, Sui ZH, Kang KH (2010) Application of SSR and AFLP to the analyses of genetic diversity in *Gracilaria lemaneiformis* (Rhodopyta). J Appl Phycol 22:607–612

Peěnikar ŽF, Buzan EV (2014) 20 years since the introduction of DNA barcoding: from theory to application. J Appl Genet 55:43–52

Perestenko LP (1980) Algae of Peter the Great Bay. Nauka, Leningrad, pp 232–239

Pires AC, Marinoni L (2010) DNA barcoding and traditional taxonomy unified through Integrative Taxonomy: a view that challenges the debate questioning both methodologies. Biota Neotrop 10:339–346

Puillandre N, Bouchet P, Boisselier-Dubayle M-C, Brisset J, Buge B, Castelin M, Chagnoux S, Christophe T, Corbari L, Lambourdière J, Lozouet P, Marani G, Rivasseau A, Silva N, Terryn Y, Tillier S, Utge J, Samadi S (2012) New taxonomy and old collections: integrating DNA barcoding into the collection curation process. Mol Ecol Resour 12:396–402

Rebours C, Marinho-Soriano E, Zertuche-González JA, Hayashi L, Vásquez JA, Kradolfer P, Soriano G, Ugarte R, Abreu MH, Bay-Larsen I, Hovelsrud G, Rødven R, Robledo D (2014) Seaweeds: an opportunity for wealth and sustainable livelihood for coastal communities. J Appl Phycol 26:1939–1951

Rindi F, Guiry MD, López-Bautista JM (2008) Distribution, morphology, and phylogeny of Klebsormidium (Klebsormidiales, Charophyceae) in urban environments in Europe. J Phycol 44:1529–1540

Robba L, Russell SJ, Barker GL, Brodie J (2006) Assessing the use of the mitochondrial cox1 marker for use in DNA barcoding of red algae (Rhodophyta). Am J Bot 93:1101–1108

Robins JH, Hingston M, Matisoo-Smith E, Ross HA (2007) Identifying Rattus species using mitochondrial DNA. Mol Ecol Notes 7(5):717–729

Ruangchuay R, Notoya M (2007) Reproductive strategy and occurrence of gametophytes of Thai laver Porphyra vietnamensis Tanaka et Pham-Hoang Ho (Bangiales, Rhodophyta) from Songkhla Province. Kasetsart J (Nat Sci) 41:143–152

Saunders GW (2005) Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future application. Philos Trans R Soc B 360:1879–1888

Saunders GW (2008) A DNA barcode examination of the red algal family Dumontiaceae in Canadian waters reveals substantial cryptic species diversity. 1. The foliose Dilsea–Neodilsea complex and Weeksia. Botany 86:773–789

Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopedia of life: an introduction to DNA barcoding. Philos Trans R Soc Lond Ser B Biol Sci 360 (1462):1805–1811

Seena S, Pascoal C, Marvanová L, Cássio F (2010) DNA barcoding of fungi: a case study using ITS sequences for identifying aquatic hyphomycete species. Fungal Divers 44:77–87

Segi T (1951) Systematic study of the genus *Polysiphonia* from Japan and its vicinity. J Fac Fish Prefect Univ Mie 1:169–272

Sherwood AR (2008) Phylogeography of Asparagopsis taxiformis (Bonnemaisoniales, Rhodophyta) in the Hawaiian Islands: two mtDNA markers support three separate introductions. Phycologia 47:79–88

Sherwood AR, Presting GG (2007) Universal primers amplify a 23S rDNA plastid marker in eukaryotic algae and cyanobacteria. J Phycol 43:605–608

Shimada S, Horiguchi T, Masuda M (2000) Two new species of Gelidium (Rhodophyta, Gelidiales), *Gelidium tenuifolium* and *Gelidium koshikianum*, from Japan. Phycol Res 48:37–46

Siow RS, Teo SS, Ho WY, Shukor MYA, Phang SM, Ho CL (2012) Molecular cloning and biochemical characterization of galactose-1-phosphate uridylytransferase from *Gracilaria changii* (Rhodophta). J Phycol 48:155–162

Steel DJ, Trewick SA, Wallis GP (2000) Heteroplasmy of mitochondrial DNA in the ophiuroid *Asterobrachion constricum*. J Hered 91:146–149

Turner NJ (2003) The ethnobotany of edible seaweed (*Porphyra abbottae* and related species; Rhodophyta: Bangiales) and its use by First Nations on the Pacific coast of Canada. Can J Bot 81:283–293

Uwai S, Yotsukura N, Serisawa Y, Muraoka D, Hiraoka M, Kogame K (2006) Intraspecific genetic diversity of *Undaria pinnatifida* in Japan, based on the mitochondrial of cox3 gene and ITS1 of nrDNA. Hydrobiologia 553:345–356

Van de Velde, Knutsen SH, Usov AI, Cerezo AS (2002) 1H and 13C high resolution NMR spectroscopy of carrageenans: application in research and industry. Trends Food Sci Technol 13:73–92

Vidal R (2008) Phylogeography of the genus *Spongites* (Corallinales, Rhodophyta) from Chile. J Phycol 44:173–182

Villanueva RD, Montaño MNE, Romero JB (2008) Iota-carrageenan from a newly farmed, rare variety of eucheumoid seaweed–"endong". J Appl Phycol 21:27–30

Vis ML, Hodge JC, Necchi OJR (2008) Phylogeography of *Batrachospermum macrosporum* (Batrachospermales, Rhodophyta) from North and South America. J Phycol 44:882–888

Vis ML, Feng J, Chiasson WB, Xie SL, Stancheva R, Entwisle TJ, Wang WL (2010) Investigation of the molecular and morphological variability in *Batrachospermum arcuatum* (Batrachospermales, Rhodophyta) from geographically distant locations. Phycologia 49:545–553

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PD (2005) DNA barcoding Australia's fish species. Philos Trans R Soc Lond Ser B Biol Sci 360(1462):1847–1857

Woelkerling WJ (1990) An introduction. In: Cole KM, Sheath RG (eds) Biology of the red algae. Cambridge University Press, Cambridge, pp 1–6

Wolf MA, Sfriso A, Andreolia C, Moroa I (2011) The presence of exotic *Hypnea flexicaulis* (Rhodophyta) in the Mediterranean Sea as indicated by morphology, rbcL and cox1 analyses. Aquat Bot 95(2011):55–58

Yang YM, Kim MS (2015) Molecular analyses for identification of the Gracilariaceae (*Rhodophyta*) from the Asia–Pacific region. Genes Genomics 37:775–787

Yang EC, Kim MS, Geraldino PJL, Sahoo D, Shin JA, Boo SM (2007) Mitochondrial cox1 and plastid rbcL genes of *Gracilaria vermiculophylla* (Gracilariaceae, Rhodophyta). J Appl Phycol 20:161–168

Yang EC, Kim MS, Paul JL, Geraldino PJL, Sahoo D, Shin J-M, Boo SM (2008) Mitochondrial cox1 and plastid rbcL genes of *Gracilaria vermiculophylla* (Gracilariaceae, Rhodophyta). J Appl Phycol 20:161–168

Yang MY, Han EG, Kim MS (2013) Molecular identification of *Grateloupia elliptica* and *G. lanceolata* (Rhodophyta) inferred from plastid rbcL and mitochondrial COI genes sequence data. Genes Genomics 35:239–246

Yeong HY, Khalid N, Phang SM (2008) Protoplast isolation and regeneration from Gracilaria changii (Gracilariales, Rhodophyta). J Appl Phycol 20:641–651

Yow YY, Lim PE, Phang SM (2011) Genetic diversity of *Gracilaria changii* (Gracilariacea, Rhodopyta) from west coast, Peninsular Malaysia based on mitochondrial cox1 gene analysis. J Appl Phycol 23:219–226

Yow YY, Lim PE, Phang SM (2013) Assessing the use of mitochondrial cox1 gene and cox 2-3 spacer for genetic diversity study of Malaysian *Gracilaria changii* (Gracilariaceae, Rhodophyta) from Peninsular Malaysia. J Appl Phycol 25:831–838

Zhang XC, He Y, Xu D (2009) Screening microsatellite sequences from *Gracilaria lemaneiformis* and its phylogenetic analysis. Period Ocean Univ China 39:259–264

Zhao F, Liu F, Liu J, Put O, Jr A, Duan D (2008) Genetic structure analysis of natural *Sargassum muticum* (Fucales, Phaeophyta) populations using RAPD and ISSR markers. J Appl Phycol 20:191–198

Zuccarello GC, West JA (2002) Phylogeography of the *Bostrychia calliptera–B. pinnata* complex (Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear, mitochondrial and plastid DNA markers. Phycologia 41:49–60

Zuccarello GC, Buchanan J, West JA (2006) Increased sampling for inferring phylogeographic patterns in *Bostrychia radicans/B. moritziana* (Rhodomelaceae, Rhodophyta) in the eastern USA. J Phycol 42:1349–1352

DNA Databases: Promises and Limitations for Plant DNA Barcoding

Selvaraj Dhivya, Mohanasundaram Saravanan, and Ramalingam Sathishkumar

Abstract DNA barcoding provides a rapid, accurate, and standardized method for species level identification using short DNA sequences. Such a standardized identification method is useful for mapping all the species on Earth as DNA sequencing technology has become cheaper now. Using this technique, it is possible to map the Earth's immense biodiversity resources, and entries can enrich the databases (DB). Hence, in this review all the available DNA barcode databases have been reviewed for its promises and limitations. The important DNA barcode databases are International Nucleotide Sequence Database Collaboration (INSDC), Barcode of Life Database (BOLD), BioBarcode, Medicinal Materials DNA Barcode Database (MMDBD), Herbarium of Med Plants of India, and PLANTS Database. All these DBs possess a vast number of specimen entries and also help in understanding their genetic relationships and evolution. Some databases also contain a chromatogram viewer, which helps the DNA sequence analysis. These barcode database servers provide bioinformatics tools as one stop solution. It promotes the rapid acquisition of biological species and to reach global standards in the aspects of standardization, depository, management, and analysis.

Keywords Database · BOLD · Specimen · Biodiversity · MMDB · Resources

1 Introduction

DNA barcoding is a technique, which provides a quick identification of species without involving the morphological cues. It uses a relatively small, standardized DNA fragment as a tag to define or discover a species. DNA barcoding uses minor differences of nucleotides in the particular gene loci of different organisms as a key for the discrimination. The gene is sequenced to know the base pair differences and

S. Dhivya · M. Saravanan · R. Sathishkumar (✉)
Plant Genetic Engineering Laboratory, Department of Biotechnology, Bharathiar University, Coimbatore, India
e-mail: rsathish@buc.edu.in

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_11

then deposited in the barcode database, which is termed as DNA barcodes. These genetic codes could be accessed through a digital library and used to identify the unknown species. The size of DNA barcode is about 400–800 bp and used to characterize all the living organisms (Savolainen 2005). Paul Hebert's group was the first to design and use the short DNA sequences for biological identification at the University of Guelph, Canada. It emerged to describe the microorganisms; now it is being applied successfully to animals and plants. A massive online digital library of barcodes will be a standard, to which the DNA barcode sequence of an unknown sample can be matched for the identification.

The key process in DNA barcoding is identifying novel candidate gene, which can be used universally. It allows the users efficiently to distinguish the species and accelerate the species discovery (van der Sande et al. 1992). DNA barcoding uses the information of one or a few regions in the genome to recognize all the species in a genus (Lahaye et al. 2008). DNA barcoding will open up new opportunities in DNA-based investigations ranging from community phylogenetics (Webb et al. 2002) to ecological genomics (van Straalen and Roelofs 2006). Despite the lack of a universal plant barcode, taxonomists, ecologists, evolutionary biologists, and conservationists are all already envisioning the purpose of a genetic identifier to a wide set of research and practical applications (Sujeevan and Hebert 2007).

For large-scale analysis, DNA barcoding can be easily studied by comparing loci across the similar set of taxa under a selected set of PCR conditions. Thus, the statistics was taken into account between the ability to amplify a locus and the rate of divergence of that locus across a phylogenetic range of taxa (Ruzicka et al. 2009). Additionally, the sequence alignment methodologies used in the DNA barcoding technique were evaluated using the following points:

- The purpose of assurance limits to species assignment
- The use of a part of sequences in database searches
- The strength of search algorithms of sequence length variation due to insertion/deletion events and the informative nature of these mutations

The ultimate aim of DNA barcoding is to discriminate the species using an automated system, so that unexplored living organisms can be named as quickly as possible before it gets extinct. DNA barcode proved to be a promising tool to identify the species across all forms of life including animals, plants, and microbes in a rapid and reliable manner. The many of the published work used a simple distance matrix analysis, a neighbor-joining (NJ) algorithm with Kimura-2-parameters (K2P).

The identification and characterization of molecular entities are the main goal in DNA barcoding studies. An ideal DNA barcode should possess the following features:

- High interspecific and low intraspecific sequence divergence
- Undergo universal amplification with standard primers
- Technically simple to analyze
- Short enough to sequence in one reaction
- Easily alignable (few insertions/deletions)

- Readily recoverable from the museum or herbarium samples and other degraded samples

1.1 Plant DNA Barcoding

DNA barcoding has a huge role in the conservation biology especially in assessing biodiversity hotspots and also monitoring the international trade of the rare species apart from the routine identification. Plants have not been given much importance in the early stages of DNA barcoding due to inability of cytochrome oxidase (*COX1*) to work as a barcode (Cho et al. 2004). The contest was set among the botanists to find a more suitable marker (Pennisi 2007). Many candidate gene regions have been recommended as possible barcodes for plants. The lack of consensus region in plants as in the case of *COX1* in animals as a universal barcode for plants has not been found till now; instead, several groups have put forward different barcode candidates successfully for the smaller taxa (Zhang et al. 2000). Several factors are considered in selecting a plant DNA barcode, like (1) universal PCR condition, (2) range of taxonomic diversity, (3) power of species differentiation, and (4) dry lab analysis and application. Approaches and application of DNA barcoding are illustrated in Fig. 1.

DNA barcoding mirrors the distribution of intra- and interspecific variation that is separated by a distance called "DNA barcoding gap" (Hebert et al. 2003; Meyer and Paulay 2005). The Consortium of Barcode of Life coordinates DNA barcoding development and implementation universally. DNA barcoding is very essential for the molecular identification of already described species (Hebert et al. 2003) and the discovery of new species (Valentini et al. 2006). The strength of DNA barcoding is directly related to the data available in the barcode libraries, which helps in building a very complete DNA barcoding database (Ekrem et al. 2007).

2 DNA Barcode Database

One of the most important components of the barcode initiative is the construction of a public reference library of species identifiers, which could be used to assign unknown specimens to known species. There are currently three main barcode databases: the International Nucleotide Sequence Database Collaborative (INSDC) which is a partnership among GenBank in the U.S., the Nucleotide Sequence Database of the European Molecular Biology Lab in Europe, and the DNA Data Bank of Japan. They have agreed to CBOL's data standards for barcode records.

Barcode of Life Database (BOLD) was created and is maintained by the University of Guelph in Ontario. It offers researchers a way to collect, manage, and analyze DNA barcode data. There are many specialized DNA barcoding databases available, which is summarized in Table 1.

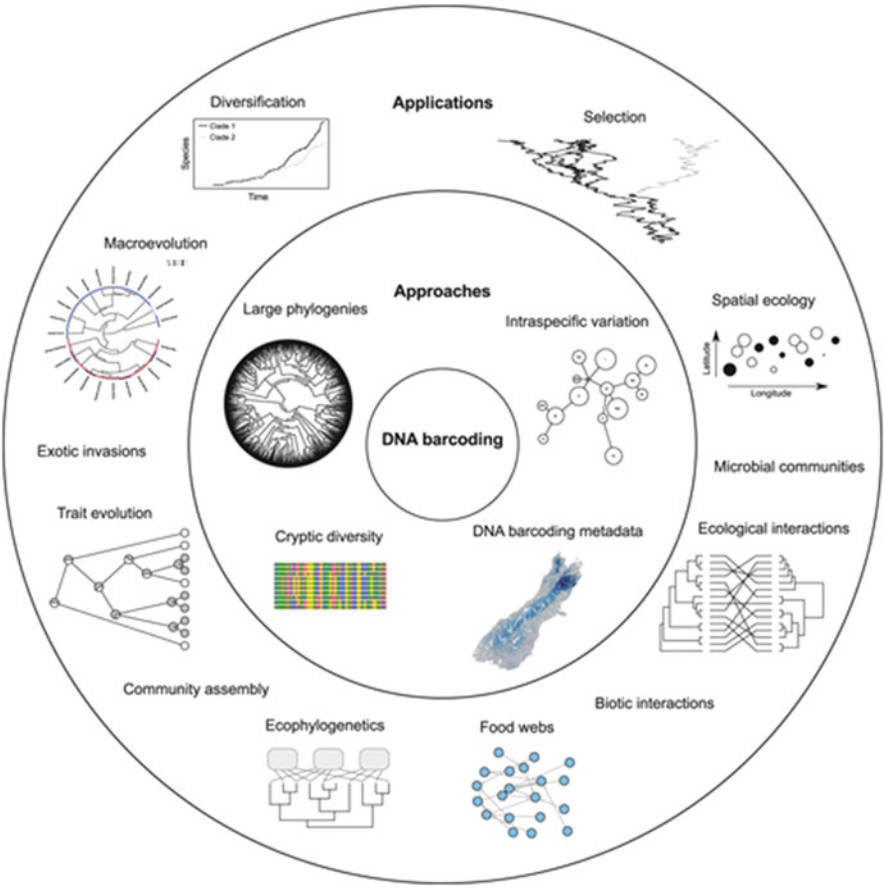


Fig. 1 Graphical representation of potential ecological application of DNA barcoding in the context of approaches and disciplines to DNA barcoding data (Joly et al. 2014)

2.1 The International Nucleotide Sequence Database Collaboration (INSDC)

The International Nucleotide Sequence Databases Collaboration has led to many beneficial projects that promise to proliferate in the molecular biology community. GenBank, along with partners DDBJ and ENA, has launched INDS (www.insdc.org). It represents the aims and policies in gathering and publishing nucleotide sequence annotations and links to the three partners' data submission and retrieval tools. Currently, the following projects are part of the collaborative effort among the three databases.

Table 1 DNA barcoding database and their web links

Name of the database	Function	Links
BOLD Systems	It applies and generates and DNA barcode data for all the species	http://www.boldsystems.org/
Medicinal material DNA barcode database (MMDB)	To identify the medicinal materials from natural sources	http://www.cuhk.edu.hk/icm/mmdbd.htm
FISHBOL	Global effort to coordinate reference sequence library for all fish species	www.fishbol.org
The Fish Barcode Information System (FBIS)	Indian subcontinent has been conceived and developed by National Bureau of Fish Genetic Resources	http://mail.nbfgr.res.in/fbis/
Plant DNA Barcode Project	Plant Barcode Database keeps track of samples from the field to final barcode sequence	http://botany.si.edu/projects/DNAbarcode/proj_db.htm
BioBarcode	DNA barcoding database and server platform for Asian biodiversity resources	http://www.asianbarcode.org
Royal Botanic Garden Edinburgh—DNA barcoding	To speed up the process of cataloguing life on Earth	http://www.rbge.org.uk/science/genetics-and-conservation/molecular-ecology/dna-barcoding
CORDIS	DNA barcoding database of plant pathogens	http://cordis.europa.eu/result/rcn/89996_en.html
Kadoorie Farm and Botanic Garden (KFBG)	DNA Barcoding of Focal Plant Taxa	www.kfbg.org/eng/dna-barcoding-of-focal-plant-taxa.aspx
Botanic Gardens Conservation International	To conserve plants, their habitats, and ecosystems	https://www.bgci.org/resources/article/0702/
Ipbes	German Barcode of Life Project	http://catalog.ipbes.net/assessments/145
QBOL	To identify quarantine organisms in support of plant health	http://www.qbol.org/en/qbol.htm
ArBOL	Database of chili pepper varieties	www.premierepos.co.uk

2.1.1 The Taxonomy Project

The goal of the collaborators is to use a unified taxonomy across all databases, largely one based on sequence information. The taxonomy project was set up as a tool for biologists worldwide and also as a shared instrument for the collaborators. This is one of the important resources used for the maintenance of genetic codes, important for the correct translation of coding sequences.

2.1.2 The Feature Table

The feature table documentation represents the shared rules that allow the three databases to exchange data on a daily basis. It represents the vocabulary that is used to describe the DNA sequence annotations as well as that of the protein sequence they encode.

2.1.3 The db_xref Qualifier

The /db_xref qualifier allows the nucleotide databases to explicit reference-specific sequences (protein sequences) or other identifiers within other databases.

2.1.4 The Country Qualifier

The country/ country qualifier indicates the country of origin of a DNA sample. This qualifier uses a controlled vocabulary and format.

3 BOLD

The Barcode of Life Data System (BOLD) is a web platform, which provides an integrated environment for the assembly and use of DNA barcode. Initially it consists of specimens and its distributions and molecular data, and also it provides analytical tools to support their data organization, validation, visualization, and publication. Recently it brings collection of iterative improvements in modules improving data dissemination, citation, and annotation. This documentation provides details on BOLD functionality, data structures, and best practices. It explains about collection, management, and publication of barcode and ancillary data. The navigation features of BOLD database consist of taxonomy browsers, identification engine, work bench, resources, log-in user interface and references, search bars, downloads, feedback link, and page footer.

3.1 Databases

The BOLD platform is comprised of a set of integrated databases. Some of these databases like the Public Data Portal and BIN database are primary sources of data, while the publication and primer databases play a supporting role. The Public Data Portal supports queries based on taxonomy (scientific names only), geography, attribution fields (i.e., collectors and taxonomists), specimen depositories, project or dataset

codes, specimen, and sequence identifiers (sampleids, museumids, processed, etc.). Free text searches are allowed; the system will try to extract meaningful terms from the search. Search terms should be separated by a space and terms containing more than one word.

3.1.1 Barcode Index Numbers

The Barcode Index Number (BIN) system clusters sequences to produce operational taxonomic units that closely correspond to species. BINs are unique, indexed with genetically identicial taxa for different studies which reside under shared identifiers. Clustering has been parameterized by using "training data" from taxonomically diverse records collected by barcoding efforts to date. Training sets are based on established taxonomy to recognize those sequence clusters that are likely to correspond to biological species. Each novel cluster is assigned a globally unique identifier that is registered in the Barcode of Life Data System (BOLD). It includes a dendrogram of all member sequences and associated literature. This enables the generation of short descriptions and annotation of data elements.

3.1.2 Primer Databases

A comprehensive registry of primers used in the generation of barcode sequences was maintained by users of BOLD.

4 BioBarcode

BioBarcode is a DNA barcoding database and web server system. It provides a reusable barcode construction system for more specific projects. It is used by biologists who have specific species information and want to build a DNA barcode database and server. It supports the compilation, storage, analysis, and publication of high-quality DNA barcode records. For many experimental biologists, building a local DNA barcoding system is expensive and time-consuming. Therefore, BioBarcode will be useful for providing the tools needed to launch successful barcoding projects in the Asian biodiversity research community. It includes software for data management and analysis, data standards, and a data repository. To establish data standard, it adopted the guidelines from CBOL and GenBank at the National Center for Biotechnology Information (NCBI) for records to maintain formal barcode status. Furthermore, it can be used for promoting international collaboration for building an Asian biodiversity system aiming to be the Asian biodiversity database server. It consists of the following components.

4.1 System Architecture and Scheme

The database structure of BioBarcode system consists of 12 tables for the management of general information on specimens and following taxonomy, project, location, attached files like image and chromatogram, and a member management table. The lineage table provides tax_names and tax_nodes from the NCBI taxonomy database to search and navigate lineage. The table contains taxonomy information of 370,000 species and their lineages from kingdom to subspecies; therefore, users can choose taxonomy and lineage information by keyword suggestion. The tables of image and trace files, such as bc_file_img and bc_file_trace, were separated to speed up and manage attachments easily. Minimal information is required for member registration, and they accept various types of geographical information, such as addresses, latitude and longitude positions on Google map, and zip codes data, currently available only for South Korea. BioBarcode was built on SQL RDBMS 5.0 as DBMS architecture.

4.2 Data Collection and Validation

BioBarcode provides users with a pathway for the direct submission of their data, which can include the information related to specimens, sequences, trace files, and images. Once data are submitted, the information was updated and the functions were directly accessible from the sequence and specimen pages. Projects were created by any registered user, a "project manager" will be subject to optional security measurement, and all data records will remain private to a single researcher or to a group of collaborators. It is represented in Fig. 2. It employs a function to assess data quality and all submitted DNA sequences, which possess experimental steps such as sample collection, DNA extraction, amplification, sequencing, translation into proteins, and also comparison against already known proteins for verification. The main target database is GenBank, and any in-house database can be used to compare with the input sequence. Sequences that pass this check are then examined for stop codons. If any potential errors are detected, the sequence is flagged. Based on these results, the blue icon shows the presence of low-quality bases. Similarly, the red icon appears when the automatically translated amino acid sequence contains stop codons. The final step is uploading the sequence data into the database with specimen information.

4.3 User Interface

The users can search the BioBarcode database using five entries. They are (1) main menu bar, (2) log-in system, (3) category, (4) statistics, and (5) using other links. The

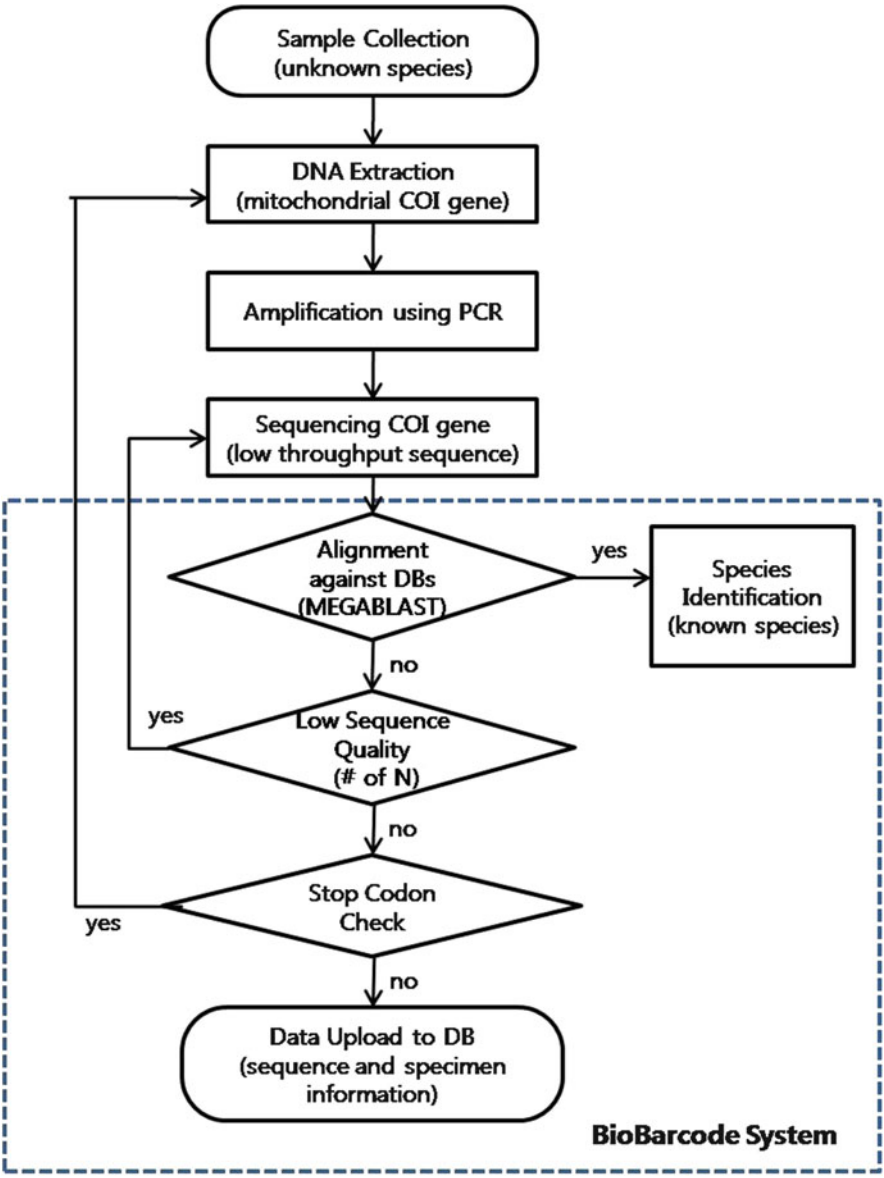


Fig. 2 Flowchart of BioBarcode for DNA sequence identification and quality validation. Explains the whole procedure from sample collection to data quality checking. The dotted line indicates how the processing occurs in BioBarcode (Jeongheui et al. 2009)

project page shows six sections such as project title, overview, geographical distribution, statistics, member list, and entries. Through the "lineage" option of the main menu section, taxonomy information based on input data which requires authorization

can be identified. Personal information and registered project lists are available in the log-in state. The statistics of registered species, specimen, sequence, projects, and trace files can be viewed on the "Statistics." The "Project" page consists of six sections like project title, general overview (e.g., number of species, specimens and sequencing reaction, dates of the start and end of project), geographical distribution, statistics, member list, and entries (e.g., barcode ID, species, and sequence with trace data).

Recently mobile interface program has become available such as Yahoo's blueprint interface http://mobile.yahoo.com/developers. They plan to implement the mobile application interface, so that researchers can easily and rapidly deposit and monitor data in real time for sampling locations, collections, and observations. Another major advantage is various Asian languages are supported by the BioBarcode DB/Server construction system.

4.4 Computational Resources

The BioBarcode system was developed with open-source software. PHP and JavaScript were used for most web pages and the main system so that any future developers can join or develop their own based on BioBarcode. mySQL was used for database management, and MEGABLAST was used to search the sequence identity. It does not require much computing resources, so stand-alone PC workstation is enough to run the server. This server is currently optimized for IE7 Internet browser. Therefore, some JavaScript or AJAX functions may not work correctly with other Internet browsers.

It is an exemplary server, which aims to provide a platform for biological researchers who want to establish their own DNA barcode database and web server system compatible with international standards. Its target are Asian researchers who have many local biodiversity resources. It is also easy to run and maintain with inexpensive open-source software (Jeongheui et al. 2009).

5 Medicinal Materials DNA Barcode Database (MMDBD)

MMDBD is hosted and maintained by The Chinese University of Hong Kong (CUHK), Republic of China. There are 1,00,000,000 sequences deposited that are mined from the other common DNA databases, which is currently available. It's mainly constructed for the identification of herbal materials and its respective products.

Specific DNA barcode database has been constructed to serve specific groups of organism. Currently there are 1661 species with 51,375 sequences available in the database, which includes China's medicinal plant species, American herbal pharmacopoeia, folk medicines, substituents, adulterants, and closely related species. The common name, scientific name, endangered species, voucher number of the samples,

primer sequences, and PCR conditions for generating barcodes are available and also additional information like specimen images and dried medicinal materials, which are provided by the Chinese medicine museum of CUHK (Lou et al. 2010).

5.1 Sequence Information in MMDBD

Considering the biodiversity richness in the world, it is widely believed that a single set of DNA barcode should be adopted. The primary goal of MMDBD is to generate standard DNA barcodes of the herbal materials, which will increase the accuracy for differentiating closely related species. MMDBD provides supplementary barcodes for reference in addition to chloroplast DNA and nuclear DNA, including ITS and 5S rRNA intergenic spacer, so a combination of two or more DNA regions can be used for problematic species. Hence, all the available chloroplast DNA sequence and two nuclear gene spacers ITS and 5S rRNA intergenic spacer are included in the database. Therefore, MMDBD consists of CBOL-proposed barcodes and other useful DNA regions for the better identification of medicinal materials.

This database contains mined DNA sequence from GenBank, INSDC, Seq, and eXtensible Markup Language (XML) files. To keep genes without standard names such as ITS, ITS1, ITS-1, or ITS2 for ITS region, multiple alternative keywords will be available for easy access of these genes in the dataset. All the extracted data were imported into the MYSQL database. MMDBD has generated 531 DNA for 189 medicinal materials, including ITS, 5S rRNA intergenic spacer, chloroplast trnL, and mitochondrial cytochrome b regions.

5.2 Software Design

MMDBD was materialized on the relational database system MYSQL (Ver. 5.0.45). Search engine is similar to NCBI BLAST (Basic Local Search Alignment Tool, v2.2.17). PERL- DBI module was employed to connect MYSQL, where we can submit our query and get instant results. MMDBD also uses AJAX (Asynchronous JavaScript and XML), which provides a rich and smooth user interface. Server-side PHP scripts are used for the requests and later on were converted to JSON (JavaScript Object Notation) format for data interchange.

MMDBD was explained in three tables, wherein general information, DNA sequence, and references of the medicinal materials piled up. “Medicinal material information” stores taxonomic, medicinal, and morphological information (Table 2). The field “CITESApp” gives the information about the endangered medicinal species in the CITES. The other field “ParmacopeiaInfor” gives the identification of all listed pharmacopeia species in MMDBD. Another field “Barcode sequence” contains sequence information, which contains 51,375 sequences. Besides, voucher

Table 2 Description of the three tables in MMDBD database

Field	Description
1. Barcode sequence	
SeqId	Sequence ID in MMDBD, which consist of identifier "MMDBD" and 9 digital number
DNA region	DNA region of medicinal material
Clone number	Clone number
Sequence	DNA sequence of a specific region
Sample index	Voucher number of authentic sample
Sample origin	Location where sample was collected
Forward primer	Forward primer for PCR
Reverse primer	Reverse primer for PCR
Source	Source of the sequence Y-from MMDBD N-from NCBI
NCBI access num	NCBI accession number
MMId	Foreign key of Medicinal material Information
RefId	Foreign key of Reference table
2. Medicinal material information	
MMId	Medicinal material Id
Species name	Species name of medicinal material
Variant name	Variant name of medicinal material
Family	Family name of medicinal material
MMName	Medicinal name of medicinal material
MMType	Type of medicinal material: "P" for plant, "A" for animal, and "F" for fungus
MedicalPart	Medical part of medicinal material
Adulterant	Adulterant of medicinal material
MMPhoto	File name of medicinal material image
NChs	Stroke number of medicinal material name in simplified Chinese
3. Reference	
RefId	Reference ID
PMID	PubMed unique identifier
Title	Title of reference
Author	Author name of reference
Journal	Journal title, publication volume, date, and page number
Abstract	Abstract of reference
Language	Language used

numbers, sample origin, DNA regions, NCBI accession numbers, and references provide information to allow the users to trace the specific DNA barcodes (Fig. 3).

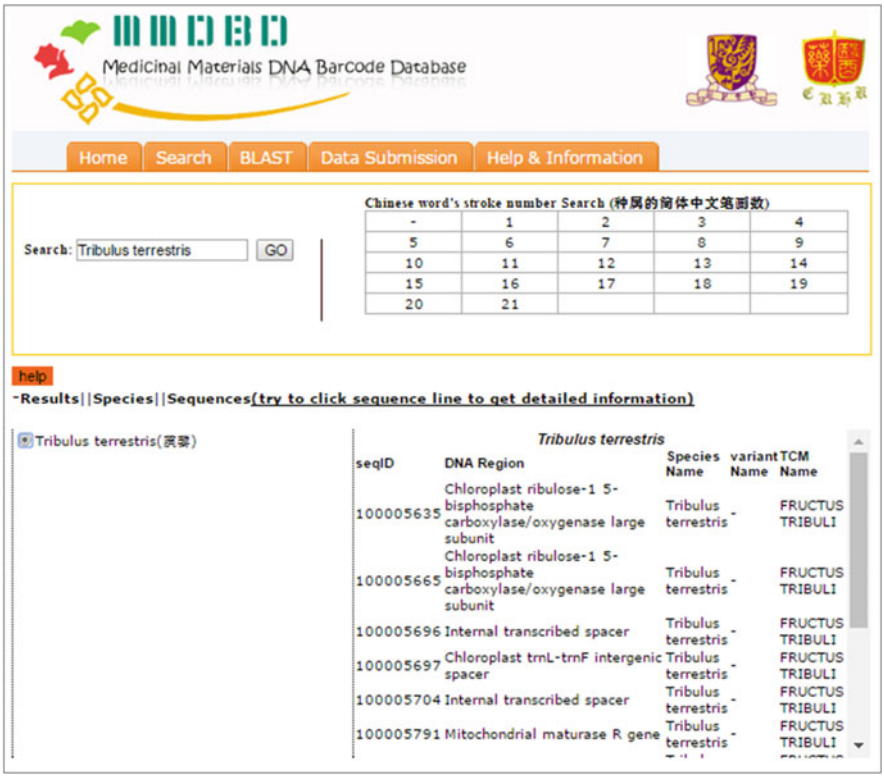

MMDBD
Medicinal Materials DNA Barcode Database
Home
Search
BLAST
Data Submission
Help & Information
Chinese word's stroke number Search (种属的简体中文笔画数)
Search: Tribulus terrestris
GO
help
-Results||Species||Sequences(try to click sequence line to get detailed information)
Tribulus terrestris(蒺藜)
Tribulus terrestris
seqID
DNA Region
Species Name
variant Name
TCM Name
100005635 Chloroplast ribulose-1 5-bisphosphate carboxylase/oxygenase large subunit Tribulus terrestris - FRUCTUS TRIBULI
100005665 Chloroplast ribulose-1 5-bisphosphate carboxylase/oxygenase large subunit Tribulus terrestris - FRUCTUS TRIBULI
100005696 Internal transcribed spacer Tribulus terrestris - FRUCTUS TRIBULI
100005697 Chloroplast trnL-trnF intergenic spacer Tribulus terrestris - FRUCTUS TRIBULI
100005704 Internal transcribed spacer Tribulus terrestris - FRUCTUS TRIBULI
100005791 Mitochondrial maturase R gene Tribulus terrestris - FRUCTUS TRIBULI

Fig. 3 Search page screenshot of Medicinal Material Database

5.3 Database Utility

MMDBD provides a simple web-based interface to retrieve the sequence information, and the URL is http://www.cuhk.edu.hk/icm/mmdbd.htm. MMDBD has three major functions, namely, database query (text-based interface), sequence similarity search, and data submission. The database is enriched by adding more number of DNA sequences from the researchers. For data submission, there is an EXCEL format template file, which facilitates users to upload batch submission of the sequences and information.

The search is executed by herb name, species name, family name, and reference of the medicinal materials. BLASTN algorithm is used for sequence similarity search. This function allows the users to search homology between their sequence of interest and the data in the database. Users are able to adjust the E-value to optimize the search results.

MMDBD provides open forum for further addition of medicinal materials. The integrated database provides users with easy access and retrieval of the data for sequence comparison. MMDBD plays an important role in authentication and

quality control of medicinal materials and benefits the herbal industry. It also helps in conservation, forensic, and systematic analysis.

6 Foundation for Revitalization of Local Health Traditions (FRLHT)

FRLHT Herbarium of Med Plants of India is a comprehensive database on Indian medicinal plants. It covers various categories of natural resources used in Indian system of medicine such as botanical and vernacular names correlation, geographical distribution, maps, propagation, trade information, etc. It is enriched based on the authentic scientific publications available in the medicinal plant research. This database contains 7637 botanical names (6198 medicinal plant species) with 1,19,183 vernacular names from 12 different languages across India. Nearly 2304 plant digital images were also available in the database.

7 Plants Database

The Plants Database is hosted and maintained by the United States Department of Agriculture (USDA) and Natural Resources Conservation Service (NRCS). Plants Database provides standardized information about the vascular plants, mosses, liverworts, hornworts, and lichens of the USA and its territories. Interactive keys and plant character datasets for selected group of US plants are available for use and testing online for the specific plants identification application. For plant identification, Plants Database provides two choices at each step (a dichotomous key); these keys select multiple characters simultaneously (polyclave key). This makes more character data available in the key and minimizes the number of steps it takes to identify a plant.

The datasets were developed cooperatively by the collaborators cited in the keys and the USDA-NRCS national plant data team. The automated plant keys run in SLIKS freeware written in JavaScript and identify specific groups of plants known to exist in the USA. Data development for the grass and legume keys was partially funded by the USDA-NRCS Agricultural Wildlife Conservation Center, and the USDA-NRCS Louisiana State Office contributed to grass key development.

8 Promises

In evaluating the promise and pitfalls of DNA barcoding, there are two areas of application in molecular diagnostics of individuals relative to described taxa and DNA-led discovery of new species. Both are inherently phylogenetic and rely on a solid taxonomic foundation, including adequate sampling of variation within species. DNA barcoding is used for large-scale screening of one or a few reference genes in order to (1) assign unknown individuals to species and (2) enhance discovery of new species (Hebert et al. 2003; Stoeckle 2009). Proponents envisage development of a comprehensive database of sequences, preferably associated with voucher specimens representing described species, against which sequences from sampled individuals can be compared. Given the long history of use of molecular markers (e.g., allozymes, rDNA, and mtDNA) for these purposes (Avise 2004), there is nothing fundamentally new in the DNA barcoding concept, except increased scale and proposed standardization. Standardization, i.e., the selection of one or more reference genes, is of proven value in the microbial community and in stimulating large-scale phylogenetic analyses, but whether "one gene fits all" is open to debate. In an effort to standardize the approach to species identification, using molecular techniques has been proposed for many species possible to characterize the same genetic marker (Blaxter 2004). There is little doubt that large-scale and standardized sequencing, when integrated with existing taxonomic practice, can contribute significantly to the challenges of identifying individuals and increasing the rate of discovering biological diversity. The approach was determined and applied for dicovering the boundary conditions. The real challenge lies with tropical taxa and with limited dispersal and thus substantial phylogeographic structure. Such analyses need to be taxonomically broad and need to extend beyond the focal geographic region to ensure that potential sister taxa are evaluated and can be discriminated. There is also the need to examine groups with frequent (possibly cryptic) hybridization, recent radiations, and high rates of gene transfer from mtDNA to the nucleus (Moritz and Cicero 2004).

9 Limitations

Genomic approach of using different DNA barcodes to deal with the existing uncertanities of ecology, phylogny and taxonomy. For characterizing the ecosystem biodiversity, databases relating the genetic information to taxa might be used as references and new taxa discovered in ecological surveys (Simon et al. 2013). The efficiency of all these approaches is currently limited by the completion of taxonomic reference database and by the choice of standard barcode with sufficient taxonomic coverage. Selecting a suitable barcode candidate makes a compromise between technical constraints over the targeted range of diversity. Most of the sequence data are mined from GenBank in which the research efforts are strongly biased toward

model organism. The barcoding of life project aims at standardizing the markers across taxa at the species level. These databases have been and are being developed fair independently, and there is a lack of general documentation operation in annotation and data exchange. Most of the databases are developed and updated using public efforts carried out by biologists and software developers in academic sectors, and it uses more efforts to share their development experiences via conferences and publications. Digital images available in the databases will provide more accurate species identification. To access the available data in the database through host and query languages with the scientific and vernacular names of plants is possible. Mainly in medicinal plant authentication, the databases play a crucial role in the identification of plant species in the admixture, and also possible adulterant sequence and taxonomical information give us the idea about the unknown species. It provides a list of diversified plant species list to compare with our plant of interest with more accuracy. The majority of papers on database explain about mostly the content and user functionality available in the database and other little information on the design and implementation of the software. An organized and comprehensive sequence information which results from the genomic research activities helps to enrich the database. Reduced data redundancy and increased consistency are achieved by larger set of sequence information in the database. Greater data integrity and independence from programs lead to the development of a user-friendly database for better understanding of the biological information. It is clear that these technical advancements and communications will become a stepping stone for the establishment of an intellect of scientific community among all types of biological database projects.

Acknowledgments The authors are gratefully acknowledging Bharathiar University, UGC-SAP, UGC WMPDF, DST-FIST, and DBT-Twinning NER for the financial support.

References

Avise JC (2004) Molecular markers, natural history and evolution, 2nd edn. Sinauer Associates, Sunderland (Massachusetts), p 684

Blaxter ML (2004) The promise of a DNA taxonomy. Philos Trans R Soc Lond Ser B Biol Sci 359 (1444):669–679. https://doi.org/10.1098/rstb.2003.1447

Cho Y, Mower JP, Qiu YL, Palmer JD (2004) Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. Proc Natl Acad Sci USA 101:17741–17746

Ekrem T, Willassen E, Stur E (2007) A comprehensive DNA sequence library is essential for identification with DNA barcodes. Mol Phylogenet Evol 43:530–542

Hebert PDN, Ratnasingham S, DeWaard JR (2003) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc R Soc Biol Sci 270:96–99

Jeongheui L, Sang-Yoon K, Sungmin K, Eo H-S, Chang-Bae K, Woon Kee P, Won K, Jong B (2009) Proceedings BioBarcode: a general DNA barcoding database and server platform for Asian biodiversity resources. BMC Genomics 10(Suppl 3):S8

Joly S, Davies TJ, Archambault A, Bruneau A, Derry A, Kembel SW, Peres-Neto P, Vamosi J, Wheeler TA (2014) Ecology in the age of DNA barcoding: the resource, the promise and the challenges ahead. Mol Ecol Resour 14(2):221–232. https://doi.org/10.1111/1755-0998.12173

Lahaye R, van der Bank M, Bogarin D, Warner J, Pupulin F (2008) DNA barcoding the floras of biodiversity hotspots. Proc Natl Acad Sci USA 105:2923–2928

Lou SK, Wong KL, Li M, Hay-But PP, Wing Tsui SK, Shaw PC (2010) An integrated web medicinal materials DNA database: MMDBD (medicinal materials DNA barcode database). BMC Genomics 11:402

Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. PLoS Biol 3:2229–2238

Moritz C, Cicero C (2004) DNA barcoding: promise and pitfalls. PLoS Biol 2(10):e354

Pennisi E (2007) Wanted: a barcode for plants. Science 318:190–191

Ruzicka J, Lukas B, Merza L, Göhler I, Abel G, Popp M, Novak J (2009) Identification of verbena officinalis based on ITS sequence analysis and RAPD-derived molecular markers. Planta Med 75:1271–1276

Savolainen V (2005) Land plants and DNA barcodes: short-term and long-term goals. Philos Trans R Soc Lond 360:1889–1895

Simon JT, Jonathan D, Annie A, Anne B, Alison D, Steven WK, Pedro PN, Jana V, Terry AW (2013) Ecology in the age of DNA barcoding: the resource, the promise and the challenges ahead. Mol Ecol Resour. https://doi.org/10.1111/1755-0998.12173

Stoeckle M (2009) Taxonomy, DNA and the barcode of life. Bioscience 53:796–797. https://doi.org/10.1641/0006-3568(2003)053[0796:TDATBC]2.0.CO;2

Sujeevan R, Hebert PDN (2007) Bold: the barcode of life data system (http://www.barcodinglife.org). Mol Ecol Notes 7(3):355–364

Valentini AS, Mattiucci P, Bondanelli SC, Webb AA, Mignucci-Giannone MM, Llavina C, Nascetti G (2006) Genetic relationships among *Anisaki*s species (Nematoda: Anisakidae) inferred from mitochondrial *COX*2 sequences and comparison with allozyme data. J Parasitol 92:156–166

van der Sande CAFM, Kwa M, van Nues RW, van Heerikhuizan H, Raue HA, Planta RJ (1992) Functional analysis of internal transcribed spacer2 of *Saccharomyces cerevisiae* ribosomal DNA. J Mol Biol 223:899–910

Van Straalen NM, Roelofs D (2006) An introduction to ecological genomics. Oxford University Press, Oxford, p 85

Webb CO, Ackerly DD, McPeek MA, Donoghue MJ (2002) Phylogenies and community ecology. Annu Rev Ecol Syst 33:475–505

Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. J Comput Biol 7(1–2):203–214

Aquatic Plant Biodiversity and DNA Barcoding

Sufia Irfan and Aishah Alatawi

Abstract Life on Earth is abundant, diverse, and intricate in organization with different interconnected strata. This complexity of life system is completely under control of genetic, organismal, and ecological diversity. Biodiversity is collection of all species of animals, plants, fungi, and microbial organisms living on Earth and the variety of habitats inhabited by them. It is the degree of nature's variety or variation within the natural system, both in terms of number and frequency. Alterations and interference by humans within the ecosystems have resulted in extensive impairment and fragmentations of habitats. Diminishing populations gives rise to inbreeding, genetic drift, and a loss of genetic variation and, therefore, creates vulnerability of species toward environmental changes. Biodiversity in aquatic ecosystem is immense. Since the beginning of time, every cradle of life and civilization has developed in and around aquatic bodies. Whether flowing or static, aquatic bodies are important to our physical chemical and biological world. A significant number of diverse plant species is found in and around aquatic bodies: flowering plant, mosses, liverworts, species of encrusting lichens, stoneworts, and other large algal species; aquatic bodies downstream support a rich diverse species of plants. Identification of rich submerged biota in freshwater and marine aquatic ecosystem is a difficult task to accomplish. In order to understand distributions, patterns, and abundances for populations or species, the collection (detection) and identification of individuals from their physical origins must be undertaken. However, species detection is sometimes extremely difficult especially in marine and aquatic environments where organisms have complex life cycles, and direct observation of early development stages is almost impossible. DNA barcode is used by biologists and ecologist to a standardized short sequence of DNA that can be recovered and featured as a unique identification marker for all species in the biosphere. DNA barcoding has presented the conventional taxonomy in digital format. This review describes a summary of available information on historical background of DNA barcoding, biodiversity in aquatic habitats, and available resources of DNA barcoding in marine

S. Irfan (✉) · A. Alatawi
Biology Department, College of Science, Faculty of Science, University of Tabuk, Tabuk, Saudi Arabia
e-mail: sirfan@ut.edu.sa

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_12

and freshwater systems. DNA barcoding has been established as a mature field of biodiversity sciences filing the conceptual gap between traditional taxonomy and different fields of molecular systematics.

Keywords Aquatic ecosystem · Biodiversity · eDNA

1 Introduction

Biodiversity or biological diversity is the sum of all the different species of animals, plants, fungi, and microbial organisms living on Earth and the variety of habitats in which they live. It is the degree of nature's variety or variation within the natural system, both in terms of number and frequency (Mutia 2009). Ecological, economic, and social values of biodiversity are all vital to economic growth, job creation, and technical development. Fishing and seafood industries, for example, are largely dependent on natural ecological systems for productivity.

Since biodiversity is based on biological resources, therefore biodiversity is considered at three major levels: genetic diversity, organismal (species) diversity, and ecosystem diversity (Fig. 1).

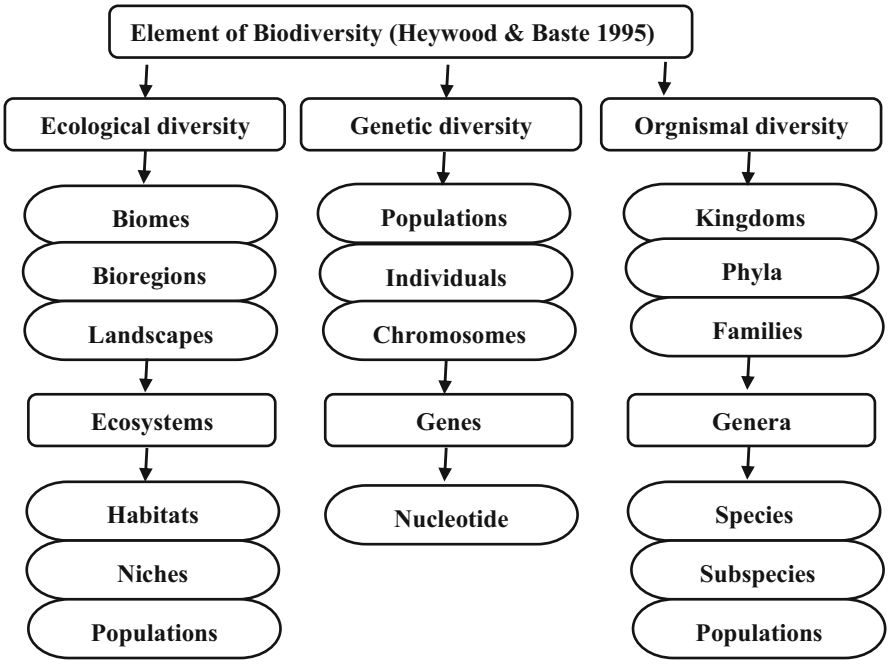


Fig. 1 Table showing different levels of biodiversity

2 Types of Biological Diversity

2.1 Species Diversity

There are 1.8 million species as described by science which are incredibly diverse. Most people are generally familiar with multicellular organisms such as plants and animals. However, the greatest metabolic diversity is found among the prokaryotic organisms of the Eubacteria and Archaea (Kearns 2010).

The species diversity can be classified as:

Species richness—the total count/number of species in a defined area. Calculated using Mangalet index and Menhink index

Species abundance—the relative numbers among species. Can be high diversity or low diversity depending on relative abundance

Taxonomic or phylogenetic diversity—the genetic relationships between the different groups of species (Mutia 2009)

The biogeographical sense of species diversity can be described in the following ways:

- Total number of species increases logarithmically with area surveyed.
- More species are found at low than high altitudes.
- Some regions called hot spots contain high local diversity.
- Substantial number of species in certain regions are highly localized (David and Flecker Alexander 1993).

2.2 Genetic Diversity

Genes are responsible for the traits exhibited by organisms. Unique genetic variants are lost as populations of species decrease in size or go extinct. Genetic variation allows species to evolve in response to diseases, predators, parasites, pollution, and climate change (Kearns 2010).

Genetic diversity in native plant populations may be underestimated in that invasive species exhibit low genetic diversity, since submersed aquatics are clonal (Sytsma 2008). The genetic diversity is also life cycle traits that influence mating patterns and gene dispersal. This determines the spatial distribution of plant genetic diversity (Rivera-Ocasio et al. 2002).

Genetic variation may be attributed to noncoding or nonfunctional DNA called as "jumping genes" or transposons. They cause mutations because of their ability to move from one place to another in the genome. The high mutation rate of this DNA

has given rise to unique pattern—"fingerprinting" within each individual species. This unique pattern—"fingerprinting"—can be used to study issues like:

- Genetic relationship between individuals and populations
- Reproductive strategies
- Breeding populations
- Development of species and populations
- Genetic drift
- Loss of genetic variation (Rieseberg et al. 2006)

A.R. Hughes et al. (2008) compiled various definitions of genetic diversity used in various studies in Table 1.

2.3 Ecosystem Diversity

The variety of habitats, biotic communities, and ecological processes that is present in the biosphere including all the abiotic factors are characteristic feature of a region (Kearns 2010; Mutia 2009).

Aquatic plants provide habitat to a lot of aquatic vertebrate and invertebrates. Plant architecture, abundance, diversity, biomass, and growth are related to each other in sustenance of the surrounding environment (Sytsma 2008).

3 Role and Importance of Biological Diversity

Biodiversity in aquatic ecosystem is immense. Since the beginning of time, every cradle of life and civilization has developed in and around aquatic bodies—either flowing in nature or static. Flowing aquatic bodies are important to our physical chemical and biological world. The continuous water cycle in the aquatic bodies plays an important role in the influx of minerals and nutrients from high level to lowlands and eventually to the sea (David and Flecker Alexander 1993).

Ecologically biodiversity plays a vital role:

- Maintaining CO_2/O_2 balance
- Regulation of biochemical cycles
- Decomposition of waste and noxious matters
- Locoregional, micro, and macro climate with impact on the world climate
- Indicators of changes in environment
- Protection against natural calamities (Mutia 2009)

Apart from human habitation, the aquatic bodies and its surroundings primarily serve as the habitat for plants and animals. According to Gleick (1996), freshwater bodies account for 40% of global fish diversity and one quarter of global vertebrate diversity. Contributing to this is a myriad of other amphibians, aquatic reptiles, and mammals;

Table 1 Definitions of genetic diversity used in ecological studies (Palmer et al. 2000)

Type of trait	Metric of diversity	Definition
Discrete allelic states	Allelic diversity	An index of molecular genetic diversity (e.g., Shannon–Wiener diversity) that incorporates information about the average number and relative frequency of alleles per locus. Allelic diversity is typically measured using molecular markers of putatively neutral loci
	Allelic richness	The average number of alleles per locus
	Genotypic richness	The number of genotypes within a population. Genotypic richness can be measured as the number of haplotypes using molecular markers, or it can be manipulated in experiments by varying the number of clonal genotypes or subfamilies
	Heterozygosity	The average proportion of loci that carry two different alleles at a single locus within an individual. Observed heterozygosity (Ho) can be estimated with codominant molecular markers, but estimates are biased by the number of individuals sampled within a population. Expected heterozygosity (He) can be estimated with both dominant and codominant markers when assumptions are made about the mode of inheritance, as well as the size and structure of populations
	Mutational diversity and effective population size (Θ)	A measure of nucleotide diversity that provides a combined measure of effective population size (N_e) and mutation rate. Θ is typically calculated using Watterson's (1975) estimator ($\theta = 4N_e\mu$), which is equal to the expected number of segregating sites between two genotypes. Estimates of θ assume an infinite number of nucleotide sites and no recombination
	Nucleotide diversity (π)	The average number of nucleotide differences per site between two random individuals selected from a population
	Percentage of polymorphic loci	The percentage of loci that are polymorphic
Continuous traits	Coefficient of genetic variance (CV)	Genetic variance in a trait (V_G) corrected by the trait mean, calculated as $(V_G^{0.5}/\text{mean}_{\text{trait}}) \times 100\%$. Unlike genetic variance, CV is not biased by the magnitude of trait means and is arguably the best measure of genetic diversity for phenotypic traits when variance scales with the trait mean
	Genetic variance (V_G)	The variance in a phenotypic trait among individuals due to genetic differences. Genetic variance is measured using parent–offspring regressions, controlled breeding designs that allow for sibling analyses or with detailed genealogical information. Total genetic variance can be further partitioned into additive and nonadditive (dominant and epistatic) components of genetic variance. Genetic

(continued)

Table 1 (continued)

Type of trait	Metric of diversity	Definition
		variance often scales positively with mean trait values
	Heritability	The ratio of the genetic variance to the total phenotypic variance in the population. Heritability values are influenced by both genetic and environmental variance and therefore offer a poor estimate of genetic diversity

together they make up to one third of all vertebrate species confined to freshwater. The total number of species living in and around freshwater bodies varies according to authors: 100,000 out of approximately 1.75 million according to Hawksworth and Kalin-Arroyo (1995) and 50,000 to 100,000 species may live in ground water bodies according to Gibert and Deharveng (2002) (Dudgeon et al. 2006).

Hynes (1970) put forward a number of diverse plant species found in and around aquatic bodies: flowering plant, mosses, liverworts, species of encrusting lichens, stoneworts, and other large algal species. Aquatic bodies downstream support a rich diverse species of plants: cool hard bottom streams containing mostly bryophytes and large soft bottom rivers supporting angiosperms. Among microphytes the plant species which is most diverse in running water bodies is the benthic algae. They occur on all possible surfaces along the river and are intimately associated with microbes and an extracellular organic matrix called *Aufwuchs*. In shaded rivers the diatom species abound in large numbers. Large aquatic bodies in the tropic have more species diversity than those in temperate regions; in addition species richness increases rapidly in lower latitudes than higher ones (David and Flecker Alexander 1993).

Flow regimes of running aquatic bodies are important. It adds to the sustaining capability of rivers and their associated floodplain. Any alteration of flow regimes often claims to be the serious and threatens wetlands and their species diversity. Numerous authors have stated that variation in plant assemblage structure (spatial and temporal) is influenced by flooding and scouring, desiccation, substrate stability, and localized variations in water velocity, turbulence, and shear stress. Patchy distribution of aquatic plants is seen due to variations in disturbance frequency and intensity, colonization success, and growth rates. Rørslett (1988, 1989) found that increased stability of baseflow and reduction of flow variability led to excessive growths of aquatic macrophytes. Similarly seedling survival and plant growth rates are affected by changes in rates of water level fluctuation and disturbance frequency and intensity (Bunn and Arthington 2002).

Authors have propounded the value of aquatic biodiversity. To enumerate a few:

- Its direct contribution to economic productivity
- Seasonal flux of wetland conditions to support livelihoods
- Value as a storehouse of genetic information
- Value in supporting the provision of ecosystem services (Dudgeon et al. 2006)

Submersed aquatic plants act as ecosystem engineers. They are critical for fish habitat, play a significant role in lentic food webs, regulate the rate of aquatic ecosystem succession, and determine shallow lake trophic system (Sytsma 2008).

4 DNA Barcoding

Science has propounded various systems to identify the living organisms around the biosphere. In Linnaean classification and binomial nomenclature, first formal system was created in 1753 to help understanding of the living organisms around humans. DNA barcoding is a modern identification system designed to help humans preserve the ecosystem and its biodiversity at a time of constant and serious threat by anthropogenic impacts. This novel identification system is considered as a master key for species identification and has profoundly influenced the process of species discovery. The practical application of DNA barcode markers in the field of specific taxonomic classification of organisms is creating a source of invaluable information in understanding species sphere, community ecology, functional trait evolution, trophic level activity, and the conservation of biodiversity (Miller 2007). Ecological studies typically need a determination of the species associated with the ecological processes. Sampling of biodiversity data for flora and fauna applying morphological characters in the identification of field data samples requires expertise in both sampling techniques and taxonomic skills, availability of which is rarely seen within a single scientific group. The recent development of DNA barcoding method for species identification has drastically simplified this step. It is especially useful for such cases in which only the remains of the specimens are available or when the morphological stage of the studied specimens is not sufficient to make a reliable identification.

This technique is cost-effective and recent in which a short DNA sequence from a standard region of mitochondrial "CO1" gene called "barcode" is used. This powerful technique can potentially be applied in biodiversity, monitoring quality of flow regime; distinguishing the indicator species of lakes, rivers, and streams; categorizing species with harmful attributes or medicinal properties; monitoring obscure species, parasitic plants, and animals; and smuggling of endangered biota from the ecosystem and their produce and disease investigations.

The DNA barcoding system is a combination of taxonomy, genetics, and computer science into a program to provide an accurate, fast, and automatable species identification by using short and standardized gene regions as internal species tag (Hubert and Hanner 2015). In order to understand distributions, patterns, and abundances for populations or species, the collection (detection) and identification of individuals from their physical origins must be undertaken. However, species detection is sometimes extremely difficult especially in marine and aquatic environments where organisms have complex life cycles, and direct observation of early development stages is almost impossible. Species detection in these environments

has been conducted using traditional direct observation methods (visual or acoustic) (Thomsen 2012; Díaz-Ferguson and Moyer 2014).

Traditional detection methods can have logistic limitations, be time consuming, expensive, and harmful to marine life forms (Thomsen 2012).The advent of novel molecular and forensic methods have provided innovative tools for detecting marine and aquatic organisms that may circumvent the aforementioned limitations (Díaz-Ferguson and Moyer 2014).

Such a novel tool is for detection of an organism's environmental DNA (eDNA). Defined as short DNA fragments that an organism leaves behind in nonliving components of the ecosystem (i.e., water, air, or sediments), eDNA is derived from either cellular DNA present in epithelial cells released by organisms to the environment through the skin, urine, feces, or mucus or extracellular DNA that is the DNA in the environment resulting from cell death and subsequent destruction of cell structure (Díaz-Ferguson and Moyer 2014).

Since the initial proposal by Hebert and colleagues (Hubert and Hanner 2015), DNA barcoding has established as a mature field of biodiversity sciences filing the conceptual gap between traditional taxonomy and different fields of molecular systematics (Hubert and Hanner 2015). It has been developed for rapid, accurate, and automatable species identification using standardized DNA sequencing as tags. In DNA barcoding, unique nucleotide sequence patterns of small DNA fragments (400–800bp) are used as specific reference collections to identify specimens and to discover overlooked species. It is of much use in areas where species identification with morphological characters is not possible due to extensive damage or delayed expression. DNA barcoding is established in animals using cytochrome oxidase subunit I in mitochondrial DNA. However in plants mtDNA has low substitution rates and a rapidly changing gene structure, which makes using COI unsuitable for barcoding in plants (Vijayan and Tsou 2010).

Table 2 lists a number of gene sequences gathered from a number of phylogenetic studies. Both coding and noncoding sequences from chloroplast DNA along with a gene from nuclear DNA have been examined for their barcoding suitability (Vijayan and Tsou 2010). In plants, three regions of the chloroplast genome (*rbcL*, *matK*, and *trnH-psbA*) and the nuclear ribosomal ITS region have been widely used as DNA barcodes, either separately or in combination. These barcodes yield relatively high species discrimination, particularly at the regional scale, at relatively low cost and have been used to build barcode libraries for local floras to address taxonomic and ecological questions (Bell et al. 2016).

5 eDNA

eDNA is defined as short DNA fragment that an organism leaves behind in nonliving components of the ecosystem (i.e., water, air, or sediments). eDNA may be either derived from cellular level or extracellular level. Cellular eDNA is derived from

Table 2 DNA segments tested for their suitability and recommendation of DNS markers for plants Vijayan and Tsou (2010)

DNA segment tested for suitability	Proposed/recommended	References
nrITS, *atpB -rbcl*, *psbM-trnC-ycf6*, *trnH-psbA*, *trnL-F*, *trnK-rps16*, *trnV-atpE rpl36-rps8*, *ycf6-psbM*	*nrITS* and *trnH-psbA*	Sytsma (2008)
nrITS, *rbcL*, *trn-psbA* *ITS1*, *accD*, *ndhJ*, *matK*, *trnH-psbA*, *rpoB*, *rpoC1*, *ycf3* *atpF-atpH*, *atpH-atpl*, *rps15-ycf1*, *ndhG-ndhI*, *psbK-psbI petA-psbJ*, *trnH-psbA*	*nrITS* and *trnH-psbA*	Sytsma (2008)
	rbcL and *trnH-psbA*	Chase et al., 2007
	atpF-atpH+psbK-psbI	Lee et al., 2007
rpoC1, *rpoB*, *matK*, *trnH-psbA*, *nrITS, trnL-F*	*rpoC1+rpoB+matK or rpoC1+matK+trnH-psbA* **(on the basis of available merits and demerits)**	Bell et al. (2016)
nrITS, *accD*, *ndhJ*, *matk*, *trnH-psbA*, *rpoB*, *rpoC1*, *ycf*	*NrITS*	Ogram et al. (1987)
trnL,(UAA) intron	*trnL,9UAA* intron	David and Flecker Alexander (1993)
accD, *matK*, *trnH-psbA*, *rbcL*, *rpoB*, *rpoC1*, UPA	*matK* and *trnH-psbA*	Newmaster et al., 2008
matK, *trnH-psbA*, *psbK-psbL*, **atpF-atpH**	*matK or matK + trnH-psbA + psbK-psbI*	Lahaye et al., 2008
accD, *matK*, *trnH-psbA*, *rbcL*, *rpoB*, *rpoC1*, *yef5*, *ndhJ*	*matK or matK + trnH-psbA*	Díaz-Ferguson and Moyer (2014)
CoxI, *23SrDNA*, *rpoB*, *rpoC1*, *rbcL*, *matK*, *trnH-psbA atpF-atpH*, *psbK-psbI*	*rbcL*, *rpoB*, *matK*, *trnH-psbA*, *atpF-atpH*	Fazekas et al., 2008
CO1, *rpoC*, *rpoB*, *rbcL-a*, *matK*, *trnH-psbA*	*trnH-psbA + rbcL-a* [based on probability of correct identification (PCI)	Erickson et al., 2008
atpF-atpH, *rpoB*, *rpoC1*, *rbcL*, *matK*, *psbK-psbI*, *trnH-psbA*	*rbcL + rpoC1 + matK + trnH-psbA*	CBOL Plant Working Group (2009)
accD, *matK*, *ndhA*, *ndhJ*, *ndhK*, *rpl22*, *rpoB*, *rpoC1 rpoC2*, *yef2*, *yef5*, *yef9*	*matK*, *rpoB*, *rpoC1*, *ndhj*, *yef5* and *accD*	Ford et al., 2009
matK, *rbcL*, *rpoB*, *rpoC1*, *trnH-psbA*	*matK*	Starr et al., 2009
atpF-atpH, *matK*, *rbcL*, *rpoB*, *rpoC1*, *psbK-psbI*, *trnH-psbA*	*rbcL + matK*	Rivera-Ocasio et al. (2002)

metabolism by-products left by organisms in the surroundings or the urine, feces, etc. Extracellular eDNA is derived from decomposition by-products left behind after cell death or destruction of cell structure (Díaz-Ferguson and Moyer 2014).

6 History of eDNA

Ogram et al. (1987) had opined that though it is a modern method, eDNA has been utilized since the mid-1980s for the detection of bacterial communities in marine sediments. A number of authors have mentioned in their studies that eDNA methods have been employed to monitor phytoplankton blooms and assess changes in biomass of bacterial communities since the 1990s (Díaz-Ferguson and Moyer 2014). Paul et al. in 1996, put forward a classification where they described two types of eDNA: (a) eDNA found in aggregates greater than 0.2 μm associated with cells (e.g., microbial eDNA) was termed particulate DNA or P-DNA, and (b) eDNA less than this size (e.g., dissolved viral DNA) was considered dissolved DNA or D-DNA(Díaz-Ferguson and Moyer 2014).

Claes Ramel (1998) discussed biodiversity and intraspecific genetic variation. Genetic variation present within species is as important as biodiversity at the level of species. Genetic variation is helpful to an organism's adaptation to environmental changes. Many organisms need to be genetically "prepared" for the changing environment and its challenges. Among several mechanisms to maintain genetic variation, balanced polymorphism is one such mechanism. Mobile DNA elements, called as "jumping genes," do not have any discernible function but behave as parasites. They cause mutations by moving from one place in the genome to another at site of insertion and thus increased genetic variation. Asexual "horizontal" transfer may occur the mechanism of which is unknown. It is mediated by microorganisms (Rieseberg et al. 2006).

Layton (2006) and Ficetola et al. (2008) in separate works have emphasized the use of eDNA for eukaryote detection to assess sources of fecal contamination in aquatic systems. French researchers first applied eDNA methods to confirm the presence of an aquatic invasive species (*Rana catesbeiana*) from water samples in a natural lotic system (Díaz-Ferguson and Moyer 2014). Jerde, Mahon, Chadderton, and Lodge (2011) in North America and Goldberg, Pilliod, Arkle, and Waits (2011) demonstrated the efficacy of eDNA as a detection tool by applying it on the detection of silver and bighead Asian carp (*Hypophthalmichthys molitrix* and *H. nobilis*) and also for federally endangered organisms. In this study the effect of seasonality on eDNA detection of amphibians was also explored. Genetic monitoring of marine mammals by Foote et al. (2012) and estimate of marine fish biodiversity by Thomsen (2012) was done using eDNA as detection tool (Díaz-Ferguson and Moyer 2014).

E. Rivera-Ocasio, T. M. Aide, and W. O. Mcmillan (2002) studied the genetic structure of a broadly distributed wetland tropical tree, *Pterocarpus officinalis*, from eight Neotropical populations using amplified fragment length polymorphisms (AFLP). A strong geographical pattern in the distribution of AFLP variation within *P. officinalis* was found. Strong genetic differences among populations and differences in the levels of AFLP variation among the eight populations sampled were studied. It was concluded that variation within this species was due to colonization history, coupled with genetic drift within local populations (Rivera-Ocasio et al. 2002).

James P. Gibbs (2000) analyzed degree historical wetland loss of wetland mosaics for two regions of the Northeastern United States. It was concluded that increased human populations have created profound reductions in wetland density and proximity. To sustain the wetland biota, all wetlands >1 acre (0.4 ha) need protection (Fazekas et al. 2009).

Stuart E. Bunn and Angela H. Arthington (2002) examined the consequences of changing flow regimes. A literature review was made to highlight the important mechanisms that link hydrology and aquatic biodiversity. Consequences of altered flow regimes were discussed. Firstly, flow determines physical habitat in streams which in turn determines biotic composition. Secondly life history of aquatic species revolves primarily in direct response to the natural flow regimes. Thirdly viability of populations of many riverine species is dependent on natural patterns of longitudinal and lateral connectivity. Finally, success of exotic, invasive, and introduced species depends on the alteration of flow regimes of aquatic bodies (Bunn and Arthington 2002).

Philip Francis Thomsen (2012) did a study to find traces in the environment for DNA and establish environmental DNA as a tool for monitoring rare and threatened species. The study was done on a diversity of rare and threatened freshwater animals—representing amphibians, fish, mammals, insects, and crustaceans. DNA was obtained directly from small water samples of lakes, ponds, and streams. Data-driven conservation of thousands of animal species which are known to be threatened or already extinct needs to be monitored. It is a challenge owing to nonstandardized methods (Thomsen 2012).

After the initial excitement over the development DNA sequences barcoding, Rob DeSalle, Mary G. Egan, and Mark Siddall (2005) discussed two concerns around DNA barcoding in their article. First being how DNA data was to be used in the context taxonomy for all known species in the world; this they called it as DNA barcode reader problem (or barcoder problem). The second problem was the reconciliation of the taxonomic community to the directives of the "DNA barcoding" initiative. This issue was important in that the classical taxonomic approach and the DNA approach needed to be reconciled in order for the "DNA barcoding" initiative to be accepted in the community (DeSalle et al. 2005).

David Dudgeon et al. (2006) explored in their article features of freshwater habitats and the biodiversity they support and their vulnerability to human activities. Threats to global freshwater biodiversity were documented. Five topics were discussed: overexploitation, water pollution, flow modification, destruction or degradation of habitat, and invasion by exotic species (Dudgeon et al. 2006).

A. Randall Hughes et al. (2008) reviewed the ecological effects of genetic diversity. Ecological processes such as primary productivity, population recovery from disturbance, interspecific competition, community structure, and fluxes of energy and nutrients are all affected by genetic diversity (Palmer et al. 2000).

Natasha de Vere et al. (2012) presented the first national DNA barcode resource. Using DNA barcoding, 97.7% coverage for rbcL, 90.2% for matK, and a dual-locus barcode for 89.7% of the native Welsh flora were assembled. Eighty-five percent of the samples were from DNA extracted from herbarium specimens (de Vere et al. 2012).

Nicolas Hubert and Robert Hanner (2015), in their article, provided an overview of the use of DNA sequences in taxonomy, since the earliest development of molecular taxonomy until the development of DNA barcoding. They presented the differences between procedures of species identification and species delineation and highlighted how DNA barcoding proposed a new paradigm that helps promote more sustainable practices in taxonomy. DNA barcoding, as discussed by the authors, is a system designed to provide species identification by using standardized gene regions as internal species tag. Foreseen since its development as a solution to speed up the pace of species discovery, DNA barcoding has established as a mature field of biodiversity sciences filing the conceptual gap between traditional taxonomy and different fields of molecular systematics. Initially proposed as a tool for species identification, DNA barcoding has also been applied in taxonomy routines for automated species delineation. Species identification and species delineation, however, should be considered as distinct activities relying on different theoretical and methodological backgrounds (Hubert and Hanner 2015).

Edgardo E. Díaz-Ferguson and Gregory R. Moyer in 2014, in their review, provided scientists with an understanding of the methods underlying eDNA detection as well as applications, key methodological considerations, and emerging areas of interest for its use in ecology and conservation of freshwater and marine environments. Genetic material (short DNA fragments) left behind by species in nonliving components of the environment (e.g., soil, sediment, or water) is defined as environmental DNA (eDNA). This DNA has been previously described as particulate DNA and has been used to detect and describe microbial communities in marine sediments since the mid-1980s and phytoplankton communities in the water column since the early 1990. More recently, eDNA has been used to monitor invasive or endangered vertebrate and invertebrate species. While there is a steady increase in the applicability of eDNA as a monitoring tool, a variety of eDNA applications are emerging in fields such as forensics, population and community ecology, and taxonomy (Díaz-Ferguson and Moyer 2014).

Punita Parikh, Krupa Unadkat, and Padamnabhi Nagar (2015) tested the potential of the rbcL marker for the identification of aquatic plants belonging to diverse families of some wetlands of Vadodara. Combination of rbcL and matK as the standard plant barcode maximum likelihood tree analysis was also performed to evaluate the discriminatory power of the rbcL gene. The classical taxonomic classification was compared with the DNA barcode tree. Tajima's D test, substitutional matrix, and nucleotide substitution model analysis were performed. It was concluded that using rbcL gene sequences, majority of the samples, i.e., (90%), were identified at genus level, but at species level, only 10% identification was possible (Parikh et al. 2015).

Karen L. Bell et al. (2016) reviewed field of pollen DNA barcoding. Applications of this technology were discussed; existing limitations and future research developments that will improve its utility in a wide range of applications were highlighted. Pollen has many applications like assessment of plant–pollinator networks, reconstruction of ancient plant communities, product authentication, allergen monitoring, and forensics. Such applications have low taxonomic significance. Pollen DNA barcoding is an alternative. Chloroplast and nuclear barcoding markers are used which are amplified from pollen (Bell et al. 2016).

7 eDNA Applications in Ecology and Conservation

The various applications of eDNA are enumerated as follows:

1. Wildlife DNA forensics
2. Detection of cryptic and endangered species/populations
3. Detection of aquatic invasive species (AIS)
4. Biodiversity and community assessment
5. Population dynamics
6. Ecosystem health
7. Trophic interactions and dietary studies
8. Species historical patterns of distribution

Various works by authors are briefly presented below in relation to the above applications (Díaz-Ferguson and Moyer 2014):

1. Wildlife DNA forensics: Ogden (2008, 2009) suggested the need of DNA forensic tools in wildlife law enforcement and forensic identification issues at the level of individual, population, or species and provide evidence for illegal wildlife trade and traceability of illegal fishing products.
2. Detection of cryptic and endangered species/populations: Goldberg et al. (2011), Takahara et al. (2012, 2013) in separate studies claimed the use of eDNA for detecting species or populations that are at low densities (e.g., threatened or endangered taxa) or are visually evasive.
3. Detection of aquatic invasive species (AIS): Darling and Mahon (2011) have stated that eDNA helps detection and confirmation of AIS in hours or days, thus allowing quick action to minimize dispersal and settlement of the invader by managers. eDNA has become a new tool for conservation managers that seek to monitor AIS as they are molecular markers and hence specific to the target species. In addition origin of the introduction and possible routes of invasion can also be determined.
4. Biodiversity and community structure: Taberlet et al. (2012) and Ji et al. (2013) have stated that multispecies identification using eDNA samples relies on next-generation sequencing and creation of taxonomic reference libraries. Metabarcoding can be used to produce biodiversity estimates that are taxonomically comprehensive, quicker to produce, and less reliant on taxonomic expertise.
5. Population dynamics: Takahara et al. (2012, 2013) correlated eDNA concentration with distributions of organisms and biomass in freshwater systems. eDNA is used for indirect measurement of population attributes such as abundance, distribution, and biomass. Population distribution in amphibians, fishes, and reptile species can be detected by eDNA concentration (Mahon et al. 2012; Minamoto et al. 2012).
6. Ecosystem health: Blanc (2001), Minamoto et al. (2009) state that invasive species and pathogens (viral or fungal) can be monitored by using eDNA. Changes in species diversity, changes in community composition, and reductions

in species diversity can all be monitored using eDNA. For risk-based decision-making of natural resources or environmental impact assessments, eDNA is useful (Veldhoen et al. 2012; Wilson and Wright 2013).
7. Trophic interactions and dietary studies: eDNA fragments or metabarcoding approach allows quantifying and estimating the relationship between predator and prey as well as herbivore and plant relationships (Zarzoso-Lacoste et al. 2013).
8. Species historical patterns of distribution: ancient eDNA in sediments, ice cores, and other environmental sources can persist for long time periods mainly in cold environments with reduced exposure or absence of light as short DNA sequences (Willerslev et al. 2003).

8 eDNA in Freshwater and Marine Systems

Lodge et al. (2012) and Thomsen (2012) have stated in their studies that eDNA technique has proven successful in a number of studies comprising a range of taxa. This is so as eDNA is abundantly available in the freshwater system usually homogeneously distributed (Thomsen 2012). eDNA gets diluted and dispersed in the marine system due to the vast amount of seawater, even though there is a high density of marine life. This makes detection of eDNA less reliable than in freshwater systems. Strong tides, high salinity environments, current system, and oceanographic events compound eDNA detection. A number of studies state that eDNA studies conducted in temperate areas are limited to detection of fishes, marine mammals, and microbial communities (Kearns 2010; Thomsen 2012).

The use of eDNA in marine systems is still a challenge; eDNA degrades rapidly in seawater (less than 7 days). In addition marine and freshwater tropical environments have high surface temperatures (sometimes above 30 °C) and elevated UV radiation at sea level that hasten eDNA degradation (Barnes et al. 2014).

9 eDNA: Perspective and Future

The perspective and future of eDNA fall on three general areas as follows: (1) resolving eDNA methodological issues, (2) improving eDNA technologies, and (3) exploring new of eDNA applications.

eDNA is an effective tool. Future success will rely on resolving eDNA-related issues like optimization of molecular assays and confirmation of positive samples. Darling and Mahon (2011) stated that to limit the uncertainty associated with eDNA false positives, the identity of all positive eDNA via Sanger sequencing should be confirmed.

10 Basic Features of DNA Barcoding

Important features of DNA barcoding are based on the universality, specificity on variation, and easiness on employment.

Rob DeSalle in 2005 (DeSalle et al. 2005) highlighted the issue of how to read the organismal barcode once it is generated. Work by Herbert et al. (2003, 2004) was mentioned, which stated that distance measures were utilized to make the inference as to species designation. However shortcomings were enumerated as follows:

- Union of classical taxonomic schemes and DNA barcoding was difficult.
- Similarity scores did not give the nearest neighbor as the closest relative (Koski and Golding 2001).
- Broad overlap of inter- and intra-specific distances within DNA exists, to delineate taxa; an objective set of criteria when using distances was lacking.

To make DNA barcoding compatible with classical approaches including character-based phylogenetic analysis was suggested by Rob DeSalle (DeSalle et al. 2005).

Numerous authors have stated that DNA barcoding should be retrievable with a single primer pair and amenable to bidirectional sequencing. The power of a DNA barcode to discriminate species determines its utility (Vijayan and Tsou 2010; Parikh et al. 2015).

DNA barcoding is usually done in animals using cytochrome oxidase subunit I gene in mitochondrial DNA. This method is well established. However due to low substitution rates in plant mitochondrial DNA and rapidly changing gene content, cytochrome oxidase subunit I gene is unsuitable for barcoding in plants (Vijayan and Tsou 2010).

According to the Consortium for the Barcoding of Life (CBOL), criteria for the development of reliable barcode data was defined as (a) candidate loci should be suitable for a wide range of taxa and (b) show high variation between species (Parikh et al. 2015).

In plants hybridization and polyploidy are common making species delimitation difficult (Fazekas et al. 2009; Rieseberg et al. 2006). Mitochondrial genome is not applicable for barcoding in plants due to frequent recombination (Palmer et al. 2000) and low rate of mutation (Cowen et al. 2006; Kress et al. 2005). Further nucleotide substitution rate of mtDNA in plants is 40–100 times lower than animal mtDNA (Cho et al. 2004; Mower et al. 2007). Standard plant DNA barcodes such as sections of two plastid genes, rbcL and matK along with supplementary markers may be required for detecting plant species. This was after evaluation of several candidate loci, by the Plant Working Group (PWG) of the Consortium for the Barcoding of Life (CBOL) (Gibbs 2000; Ramel 1998). A study by Neem et al. (2014) revealed that matK sequence had species discrimination power of 66.66%, than rbcL, which was about 90% (Randall Hughes et al. 2008). However, ribulose bisphosphate carboxylase (rbcL) and maturase (matK) plastid genes are marked as sophisticated barcode genes in plants. DNA internal transcribed spacer (ITS) region is used as

complementary region (Chase et al. 2005; CBOL Plant Working Group 2009; Consortium for the Barcode of Life 2009). It has been a decade in obtaining invaluable information using DNA barcoding as a biological programming tool to better understand nature and the environment. DNA barcode library has expanded across the tree of life, ecological habitats, and phytogeography. In the coming time, new and more extensive application will be explored and developed.

11 Conclusion

For the detection of aquatic invasive species or endangered organisms, inventory and monitoring DNA in environmental samples can be collected. Despite negative criticism from a few researchers, DNA barcoding has received immense popularity in the scientific community, especially among biologists. It is opted as an important technique for applications in conservation biology and ecology. Detection of eDNA from freshwater and marine ecosystems depends on proper experimental design. Current literature suggests that among the primers developed, *rbcL* is very useful for the barcoding of plant species. Environmental DNA can be a quick, cost-effective, and standardized way to obtain basic data on distribution and abundance of species. This will enable efficient deployment of limited conservation resources and taxonomic expertise toward conservation of rare and threatened species. DNA barcoding is all set to become a global standard for species identification in applied field biology.

References

Bell KL et al (2016) Pollen DNA barcoding: current applications and future prospects. Genome 59:629–640

Bunn SE, Arthington AH (2002) Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity. Environ Manag 30(4):492–507

CBOL Plant Working Group (2009) A DNA barcode for land plants. Proc Natl Acad Sci USA 106:12794–12797

Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP, Haidar N, Savolainen V (2005) Land plants and DNA barcodes: short-term and long-term goals. Philos Trans R Soc Lond B Biol Sci 360:1889–1895

Chase MW et al (2007) A proposal for a standardized protocol to barcode all land plants. Taxon 56:295–299

Cho Y, Mower JP, Qiu YL, Palmer JD (2004) Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. Proc Natl Acad Sci U S A 101:17741–17746

Consortium for the Barcode of Life (2009) CBOL approves matK and rbcL as the BARCODE regions for land plants. doi:https://doi.org/10.3109/19401736.2015.1089572. DNA barcoding 4469. www.barcoding.si.edu/PDF/PlantWG/CBOL%20Decision%20%20Plant%20Barcode%20Regions.pdf. Accessed 15 Dec 2017

Cowen RK, Paris CB, Srinivasan A (2006) Scaling of connectivity in marine populations. Science 311:522–527

David AJ, Flecker Alexander S (1993) Biodiversity conservation in running waters. Bioscience 43 (1):32

de Vere N et al (2012) DNA barcoding the native flowering plants and conifers of Wales. PLoS One 7(6):e37945. https://doi.org/10.1371/journal.pone.0037945

DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. Philos Trans R Soc B 360:1905–1916

Díaz-Ferguson EE, Moyer GR (2014) History, applications, methodological issues and perspectives for the use of environmental DNA (eDNA) in marine and freshwater environments. Rev Biol Trop 62(4):1273–1284

Dudgeon D et al (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. Biol Rev 81:163–182

Erickson DL, Spouge J, Resch A, Weigt LA, Kress JW (2008) DNA barcoding in land plants: developing standards to quantify and maximize success. Taxon 57:1304–1316

Fazekas AJ et al (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS One 3:e2802

Fazekas AJ, Kesanakurti PR, Burgess KS, Percy DM, Graham SW, Barrett SC, Newmaster SG et al (2009) Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? Mol Ecol Resour 9:130–139

Ford CS et al (2009) Selection of candidate coding DNA barcoding regions for use on land plants. Bot Linn Soc 159:1–11

Ficetola F, Miaud C, Pompanon F, Taberlet P (2008) Species detection using environmental DNA from water samples. Biol Lett 4:423–425

Gibbs JP (2000) Wetland loss and biodiversity conservation. Conserv Biol 14(1):314–317

Hebert PDN, Cywinska A, Ball SL, de Waard JR (2003) Biological identifications through DNA barcodes. Proc R Soc Lond B 270:313–321

Hubert N, Hanner R (2015) DNA barcoding, species delineation and taxonomy: a historical perspective. DNA Barcodes 3:44–58

Heywood VH, Baste I (1995) Introduction. In: Heywood VH, Watson RT (eds) Global biodiversity assessment. Cambridge University Press, Cambridge, pp 1–19

Kearns CA (2010) Conservation of biodiversity. Nat Educ Knowl 3(10):7

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci U S A 102:8369–8374

Lahaye R, Savolainen V, Duthoit S, Maurin O, van der Bank M (2008) A test of psbK–psbI and atpF–atpH as potential plant DNA barcodes using the flora of the Kruger National Park as a model system (South Africa). Nat Proc. http://hdl.handle.net/10101/npre.2008.1896.1

Layton A (2006) Development of Bactereroides 16S rRNA gene Taqman-base real time PCR assays for estimation of total, human and bovine fecal pollution in rivers. Appl Environ Microbiol 72(6):4214–4224

Lee HL, Yi DK, Kim JS, Kim KJ (2007) Development of plant DNA barcoding markers from the variable noncoding regions of chloroplast genome. In: Abstract presented at the Second International Barcode of Life Conference, Academia Sinica, Taipei, Taiwan, September 2017, pp 18–20. http://www.bolinfonet.org/conferences/assets/files/confrence_abstract_book.pdf

Miller SE (2007) DNA barcoding and the renaissance of taxonomy. Proc Nat Acad Sci USA 104:4775–4776

Mower JP, Touzet P, Gummow JS, Delph LF, Palmer JD (2007) Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. BMC Evol Biol 7:135. https://doi.org/10.1186/1471-2148-7-135

Mutia TM (2009) Biodiversity conservation; Presented at short course IV on exploration for geothermal resources, organized by UNU-GTP, KenGen and GDC, at Lake Naivasha, Kenya, November 1–22, 2009

Newmaster SG, Fazekas AJ, Steeves RAD, Janovec J (2008) Testing candidate plant barcode regions in the Myristicaceae. Mol Ecol Notes 8:480–490

Ogram A, Sayler G, Barkay T (1987) The extraction and purification of microbial DNA from sediments. J Microbiol Methods 7:57–66

Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu YL, Song K (2000) Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. Proc Natl Acad Sci U S A 97:6960–6966

Parikh P, Unadka K, Nagar P (2015) Evolutionary analysis based on DNA barcoding of certain aquatic plants using rbcL gene sequences. Int J Sci Eng Res 6(7):1410–1421

Ramel C (1998) Biodiversity and intraspecific genetic variation. Pure Appl Chem 70 (11):2079–2084

Randall Hughes A et al (2008) Ecological consequences of genetic diversity. Ecol Lett 11:609–623

Rieseberg LH, Wood TE, Baack EJ (2006) The nature of plant species. Nature 440:524–527

Rivera-Ocasio E, Aide TM, Mcmillan WO (2002) Patterns of genetic diversity and biogeographical history of the tropical wetland tree, *Pterocarpus officinalis* (Jacq.), in the Caribbean basin. Mol Ecol 11:675–683

Starr JR, Naczi RFC, Chouinard BN (2009) Plant DNA barcodes and species resolution in sedges (Carex, Cyperaceae). Mol Ecol Resour 9(Suppl. 1):151–163

Sytsma MD (2008) Introduction: workshop on submersed aquatic plant research priorities. J Aquat Plant Manag 46:1–7

Thomsen P (2012) Monitoring endangered freshwater biodiversity using environmental DNA. Mol Ecol 21:2565–2573

Vijayan K, Tsou CH (2010) DNA barcoding in plants: taxonomy in a new perspective. Curr Sci 99 (11):1530–1541

Part IV
DNA Barcoding in Animals

DNA Barcoding of Mosquito Species

Lalita Gupta, Sanjeev Kumar, and Kuldeep Gupta

Abstract Insects are the very diverse group of animals that exist on earth since its origin. Their genomic diversity as well as existence also imparts into the species variability, which also stays alive with their future counterparts. Among insects, there are about 3500 species of mosquitoes (Diptera, Culicidae) which have been described worldwide. In this chapter, we deal with DNA barcoding, the use of short, standardized genomic segments as markers for species identification. DNA barcoding is a novel concept for taxonomic categorization of various insects species and has been applied to study a large number of eukaryotes. At present, 10% of insects species have been DNA barcoded. In insects, interspecific variation in DNA sequences of some genes is much higher than intraspecific variability. This provides an opportunity to use DNA sequences for species identification through a fragment of the nuclear or mitochondrial gene which has been selected as the standard barcoding region. Mitochondrial molecular markers always have been a good choice over the nuclear molecular marker by having the less number of evolutionary changes in their subcompartment genome. DNA barcoding is a recent and widely used molecular-based identification system that aims to identify biological specimens and assigning them to a given species. It can be considered the core of an integrated taxonomic system, where bioinformatics and phylogenetics play a key role. DNA barcoding can be used to discover cryptic, closely related and morphologically similar species that were overlooked earlier by the traditional morphology-based approaches.

Keywords Cytochrome oxidase I (COI) · Barcode of Life Data System (BOLD) · DNA amplification · Genetic diversity · Species diversity · Insect biodiversity

L. Gupta (✉)
Department of Zoology, Chaudhary Bansi Lal University, Bhiwani, Haryana, India

S. Kumar
Department of Biotechnology, Chaudhary Bansi Lal University, Bhiwani, Haryana, India

K. Gupta
Department of Biomedical Sciences, School of Medicine and Health Sciences, University of North Dakota, Grand Forks, ND, USA

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_13

1 Insect Diversity and DNA Barcode

Insects are the major group of animals categorized under phylum *Arthropoda*, which evolved with huge diversity since the ancient period of time. To catalogue the vast numbers of insect species, entomologists came up with the idea of classifying them on the basis of taxonomy, which is a branch of science that helps us to describe a living being on the basis of morphological features. In this context identification of insects require a number of specialists to categorize them. In addition, insects comprise the most variable group, with more than 1 million described species, and many more still need to be discovered (Giangrande 2003; Grimaldi and Engel 2005). They affect mankind in a numerous ways, both harmful (e.g., disease vectors, crop pests) and helpful (e.g., pollinators, producer of various commercial products, biological control agents). Research on insects has significantly added to our understanding of evolution, ecology, and the genetic control of development. However, a basic requirement in gaining useful knowledge about any group is the ability to describe, classify, and subsequently identify its taxon. Groups, such as insects, offer great challenges to the taxonomic venture simply because of their diversity. The identification of species by traditional morphological methods is complex and usually requires specialist knowledge. The recognition, accounting, and naming of new insect species are more important than the number of taxonomic specialists, whose workforce is in decline (Godfray 2002).

Due to their biggest range of archeological and living samples, the standard method of utilization of morphological characters becomes challenging because of various factors such as cryptic species or phenotypical variations. Until now, specimens were identified using morphological features like the shape, size, and color of body parts. Traditionally, an experienced professional or entomologist is needed to identifications of insect species using morphological keys. Nevertheless, the development of insects also undergoes different life stages that are morphologically dissimilar from the adults and craft a problem or misguide in identifying the new specimens.

The inherent limitations of the conventional taxonomical identification methods indicate the need for a new and simple method of taxon identification. Although there is always a risk to compare the species, only consider the small chunk of DNA sequence and draw any conclusion (Moritz and Cicero 2004). Conversely, the concept of DNA barcoding also has molecular limitations (Will and Rubinoff 2004; Will et al. 2005). For example, based on the comparison of D3 domain of 28S (28S-D3), *An. fluviatilis* species S has been considered as synonymous to the *An. minimus* species C (Garros et al. 2005; Chen et al. 2006). However, subsequent investigation of the conspecificity of these two species, based on ITS2 and D2–D3 domains of 28S rDNA regions, suggested that *An. fluviatilis* S and *An. minimus* C do not deserve to be synonymous (Singh et al. 2006).

DNA barcode is a novel tool to understand the evolutionary history and phylogenetic relationship with other insect groups in the fastest way. Thirteen years ago, since its perception, DNA barcode method is also superior method in case of

damaged or an immature stage of development; even specialists may be unable to make identifications. Barcoding solves these problems because even nonspecialists can obtain barcodes from tiny amount of tissue. This is not to say that traditional taxonomy has become less important. Rather, DNA barcoding can serve a dual purpose as a new tool in the taxonomists toolbox supplementing their knowledge as well as being an innovative device for non-experts who need to make a quick identification (Ebach and Holdrege 2005).

Practically, DNA barcoding is not such a simple tool because DNA sequences are subject to all the natural impediment of molecular evolution and can show substantial variation within species as well as population (Mallet and Willmott 2003; Meyer and Paulay 2005). To avoid the complexity in DNA barcoding, a particular genomic region can be used as an important choice. This region must be camparable and have a high rate of evolution to show the genetic variation between closely related species. On the other hand, it also must have enough conservation in the sequence to amplify the target gene from a particular set of PCR primers. If a gene with a relatively rapid evolution is chosen, it might be sufficient for species determination. The gene region used as the standard barcode for all animal groups is a ~650 base-pair region near the 5′ end of the mitochondrial cytochrome c oxidase I (COI) gene. COI gene has been proven highly effective to identify birds, butterflies, fish, flies, and many other animal groups. However, COI gene is not an effective barcode region because it evolved too slowly (Hebert et al. 2003a).

In insects, COI gene provides an ideal species identification marker due to the lack of introns, restricted exposure to recombination, simple alignment and the availability of robust primer sites. Usually, sequence variation in this region shows large interspecific, but small intraspecific, divergences, which ultimately resulted into closely related species frequently from clearly distinguishable clusters on a distance-based phylogenetic tree (Porter et al. 2014). This molecular marker is strongly in accordance with the same species which were recognized previously through morphological and behavioral characteristics (Hebert et al. 2003a, b). Important advantages of a sequence-based identification include the digital nature of a DNA sequence, which allows it to understand without prejudice. Furthermore, DNA extracts from any life stage of insects—egg, larva, or adult—or from fragments of dead material will create a similar result. In addition to that, social insects such as ants and termites exhibit highly divergent caste categories that, in some cases, have been diagnosed incorrectly as distinct species—DNA barcoding promises to overcome such ambiguities (Smith et al. 2005).

The approach of DNA-based species identification entails (a) extraction of target genomic DNA from the sample, (b) analysis of the sequencing and (c) utilization of the sequence information to distinguish at species and subspecies level. DNA barcode helps to resolve several species and cryptic species mystery. For example, it was noticed from recent studies that only 9 species were documented by morphological criterion, while DNA barcoding recovered 13 distinct species on a Turkish *Culex* faunal survey of arbovirus screening programs (Gunay et al. 2015). In another study related to zoogeographical and DNA barcode approach in a lepidopteran *Astraptes fulgerator* revealed 10 largely sympatric species in the forest of

Northwestern Costa Rica raised the fact that this insect comprises more number of hidden species (Hebert et al. 2004). Sexual dimorphism has been a source of complications for taxonomists since so long. Janzen et al. (2005) described a case where each sex of the butterfly *Saliana severus* was recorded as a separate species in an inventory, until barcoding revealed that males and females had the same COI sequences and subsequently led to the recognition of a single, highly sexually dimorphic species. Since its development in 2003, this technology has applied not only basic taxonomic identification but also in fields as biodiversity monitoring, phylogenomics, and ecosystem reconstruction, with new uses emerging in public health and agriculture (Murugan et al. 2016). DNA barcoding is applied to a large-scale biodiversity monitoring among many naturalist and taxonomist. Barcode of Life Database System (BOLD) was created and is maintained by the University of Guelph in Ontario. It offers researchers an excellent way to collect, manage, and analyze DNA barcode data. The specimens are identified by finding the closest matching references in the existing database. CBOL's Data Analysis Working Group has created the Barcode of Life Data Portal which provides the researchers a new and more flexible ways to store, manage, analyze, and display their barcode data. The advantage of DNA barcoding and its potential to hold up the data leads to the propagation of reliable taxonomic information (Singer and Hajibabaei 2009).

2 Insect Species Identification

Species have been considered as a basic unit of taxonomical hierarchy, but the issue of how best to demarcate the species remains controversial (Claridge et al. 1997). Although reproductive isolation is often considered the most important indicator of species status, varied criteria have also been employed to delimit the species, such as ecological niches, mate recognition, genetic cohesion, and evolutionary history. These diverse criteria necessarily lead to ambiguity, which can have important implications to understand the biodiversity and conservation to estimate the taxon richness (Mayden 1997). In practice, one insect species recognized by the presence of one or more apparently fixed or nonoverlapping diagnostic differences and detailed examination of genital morphology have represented the gold standard for species definition for nearly a century (Eberhard 1985). Application of this factor is hindered by the lack of an appropriate methodology to quantify shape variation and by problematic homology evaluation (Arnqvist 1998). All these factors require prior experience and collectively make species identification an extremely specialized and time-consuming science, and even expert taxonomists may have difficulty to reach the conclusion. Furthermore, this dependence on diagnostic characters that are present only in the adult life stage solemnly limits the identification, as many specimens lack these characters (Balakrishnan 2005).

With the new perception, one insect species differs from other in morphology, ecology, behavior, and DNA sequences. The use of DNA sequences to gain information about the taxonomic affinities of an unknown specimen reveal its earliest

adoption in the least morphologically tractable groups such as viruses and bacteria (Theron and Cloete 2000; Blaxter et al. 2005). This method relies on the use of algorithms such as Basic Local Alignment Search Tool (BLAST) which enables the DNA sequence comparison in conjunction with DNA databases (Altschul et al. 1990). Hence, at least in standard, a particular gene or gene fragment can be used to identify a given species in a way similar to the trade barcodes that uniquely identify each consumer product. DNA barcoding promises the ability to automate the identification of specimens by determining the sequence of some specific genomic markers, e.g., ribosomal DNA (ITS2, D2, D3, IGS) and mitochondrial DNA (COI, COII, Cyt b, ND5). With the introduction of new sequencing techniques like Illumina MiSeq platform and next-generation sequencing, it is very accurate and rapid to identify the species from single specimen with few nanograms of DNA (Shokralla et al. 2014, 2015). Major controversy of this modern techniques arose when only comparative DNA sequence analyses of some specific region of gene were used as firsthand criteria to differentiate between the status of sibling species, subspecies, and specific species within the same taxon.

3 DNA Barcodes of Insect Life

The flies (Diptera) represent very diverse insect order, with around 150,000 described species (Grimaldi and Engel 2005; Beutel and Pohl 2006). The collection of DNA barcode markers has now been applied to specific taxonomic groups of insects which is proving very useful for understanding species boundaries, functional trait of evolution, and the conservation of biodiversity. The application of next-generation sequencing (NGS) technology will greatly expand the versatility of DNA barcodes across the living world. Sequences can be produced in the laboratory from a sample within a few hours in a largely automated fashion (Ball and Armstrong 2006). Recent advancements in high-throughput DNA sequencing technology have made the generation of huge DNA data very simple (Shendure et al. 2004).

There are many international barcoding activities dedicated to the insects which have been developed for public reference using barcode sequence libraries. The International Barcode of Life (iBOL) project is the largest biodiversity genomics initiative till date. Work over the past few years has produced DNA barcode records for more than 50,000 species (Singer and Hajibabaei 2009). The iBOL project was started in 2010 and will be over by 2020. Consortium members has recorded 5.3 million DNA barcodes from 580,000 species into the interactive Barcode of Life Data (BOLD) System, to construct the DNA barcode reference library for all the possible eukaryotes present in the earth. It is a not-for-profit organization controled by an international board of directors representing funding organizations. The iBOL secretariat is housed by the Biodiversity Institute of Ontario at the University of Guelph (source: http://lepbarcoding.org/iBOLoverview). There are many barcoding projects in the form of databases and few of them are dedicated to some very important and highly prosperous insects order as discussed below.

3.1 Bee Barcode of Life Initiative (Bee-BOL)

Order Lepidoptera comprises the butterflies and moths which are the second-most diverse order of insects with about 160,000 known species and likely as many still awaiting description (Hajibabaei et al. 2006). As a major target for analysis since the earliest developments of DNA barcoding, Lepidoptera have strongly contributed to the efficiency of the method and its applicability to highly diverse lineages. The aim of this database is to build a COI barcode library for all Lepidoptera species.

This library will permit the rapid, reliable identification of Lepidoptera at any stage of their development (egg, caterpillar, pupa, or adult) and will facilitate the discovery and description of new species (Bächtold et al. 2017). From Trinidad, West Indies, six new species of butterflies has been recorded. Total 781 species of butterflies are now known in Trinidad (Cock and Alston-Smith 2017)

These barcode libraries, in combination with the ones built for other groups of terrestrial animals, will make possible the detailed biodiversity maps required to guide the positioning of protected areas and to monitor the status of terrestrial life. Bee-BOL, the global Bee Barcode of Life Initiative, coordinates the assembly of all ~20,000 bee species in the form of an electronic database into a standardized reference sequence library. This database contains DNA barcodes, images, and geospatial coordinates of examined specimens. In addition, the database also provides linkages to voucher specimens, information on species distributions, nomenclature, authoritative taxonomic information, and literature citations.

3.2 Trichoptera Barcode of Life (Trichoptera BOL)

Trichoptera Barcode of Life is a long-term project to barcode the world's approximately 13,000 species of caddisflies. Launched in mid-2007, this DNA barcoding database aims to build a comprehensive COI barcode reference library for all caddisfly (Insecta, Trichoptera) species in the world. Fulfilling such an ambitious goal is not an easy task. This project is expected to be a long-term research program. Interestingly, the Trichoptera Barcode of Life operation focuses on a standardized mitochondrial COI gene fragment instead of trying to explore multiple gene markers across a variety of taxa.

3.3 Tephritid Barcode Initiative (TBI)

The Tephritidae are one of two fly families referred to as fruit flies, the other family being the Drosophilidae. The Tephritidae family does not include the biological model organisms of the genus *Drosophila*, which is often called the "common fruit fly." Nearly 5000 described species of tephritid fruit fly are categorized in almost

500 genera of the Tephritidae. Description, recategorization and genetic analyses are constantly changing the taxonomy of this family. To distinguish them from the Drosophilidae, the Tephritidae are sometimes called peacock flies, in reference to their elaborate and colorful markings.

TBI, the Tephritid Barcode Initiative, is a 2-year "demonstration project" that will create an operational system for identifying fruit flies around the world. TBI will barcode at least four representatives of all tephritid fruit flies that are either (1) agricultural pests, (2) beneficial species used for biological control of other pests, (3) closely related to pests or beneficial species, and (4) representative species from other families of tephritids. TBI plans to obtain barcodes from approximately 2000 species of the estimated 4500 known tephritid species.

3.4 Mosquito Barcode Initiative (MBI)

Mosquito (Diptera: Culicidae) are the most important contributor of causing nuisance and transmitting deadly pathogens to human. In 1897, Ronald Ross, a Scottish physician working in India, discovered that only members of one mosquito genus, *Anopheles*, carry the *Plasmodium* parasite, the single-celled organism that causes malaria in humans. This disclosure reflected into difficult efforts, involving the dissection of vast number of mosquitoes and looking into the stomachs which carry the parasites. A century later, multiple tools are available to identify the accurate and rapid characterization of the disease-causing vector species. Many molecular diagnostic tools were functional before the DNA barcoding which include allozyme electrophoresis (Green et al. 1992), DNA hybridization (Beebe et al. 1996) and restriction fragment length polymorphism (RFLP) (Fanello et al. 2002) to identify the mosquitoes. Sequencing-based advancements have also been extensively utilized, although many of them focussed on nuclear genes other than mitochondrial COI gene, like 28S ribosomal unit and ITS-2 (Kent et al. 2004; Marrelli et al. 2005). Recently, a number of studies have shown that the customary COI barcode marker also serves efficiently for species-level discrimination in surveys of Canadian (Cywinska et al. 2006), Chinese (Wang et al. 2012), and Indian (Kumar et al. 2007; Daravath et al. 2015; Gupta et al. 2016) mosquitoes. Foley et al. (2007) constructed a molecular phylogeny of the Australian *Anopheles annulipes* species complex based on four different loci, both nuclear and mitochondrial (COI, COII, ITS2 and EF-1) and found that 11 of the 17 sibling species (65%) had unique COI sequences which holds some assurance for diagnosing species within the complex (Foley et al. 2007).

The Mosquito Barcode Initiative (MBI) is an another demonstration project aimed at producing a global operational system for identifying mosquitoes. MBI barcoded at least five specimens from 80% of the 3200 mosquito species. Disease-bearing species and their closest relatives will be the highest priority. Among insects, dipterans have the greatest negative impact on human health and livestock, with groups such as mosquitoes and tse-tse fly acting as vectors of the agents for several major diseases including malaria, sleeping sickness and filariasis. Table 1 narrows down the numbers

Table 1 Clinically important disease vectors species with DNA barcode (source: http://www.barcodinglife.com/index.php/TaxBrowser_Home)

	Specimen records	Specimens with barcodes	Species with barcodes
Phylum *Arthropoda*	5,467,603	4,262,392	183,363
Class *Insecta*	4,866,998	3,848,360	165,499
Order *Diptera*	1,863,706	1,646,132	20,103
Family *Culicidae*	59,847	41,932	1066
Subfamily *Anophelinae*	16,501	10,754	358
Genus *Anopheles*	16,476	10,733	354
Subfamily *Culicinae*	41,398	1145	708
Genus *Aedes*	17,653	15,715	146
Genus *Culex*	13,797	9055	198

of dipterans and then into the mosquito species barcoded so far among other important insect species. Correct identification and categorization of vector species of mosquito is the central part to the study of vector-borne disease surveillance and control. The conventional morphology-based identification methods are time-consuming and not always sufficient to identify at the species level. Therefore, a multidisciplinary approach, including morphological and molecular techniques is essential for mosquito species identification.

4 Conclusion

From its inception in 2003, the primary goal of this technology was species identification which is growning day by day. As a research tool for taxonomist, it is now expanding from basic taxonomic identification to biodiversity monitoring and ecosystem reconstruction. The new emerging uses of this innovative techniques are in public health, agriculture and biodiversity management. The principle relies on specimen identification using a partial sequence for COI gene. Investigators can identify specimen by first extracting its DNA and then amplifying and sequencing of COI gene and then compare the sequence from the COI sequences for all known species. The use of DNA sequences for species identification predates the formal proposal of DNA barcoding which aims to provide a new identification tool for unidentified specimens or cryptic diversity in order Diptera. The vision of a handheld instant species identification device remains an exploratory idea for the future to gather barcode data of life (Savolainen et al. 2005; Hebert and Gregory 2005). Through this approach, species diagnosis for any insects in the world using DNA sequences will soon become a reality.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Arnqvist G (1998) Comparative evidence for the evolution of genitalia by sexual selection. Nature 393:784–786

Bächtold A, Kaminski LA, Magaldi LM, Oliveira PS, Del-Claro K, Janzen DH, Burns JM, Grishin N, Hajibabaei M, Hallwachs W et al (2017) Integrative data helps the assessment of a butterfly within the *Udranomia kikkawai* species complex (Lepidoptera: Hesperiidae): immature stages, natural history, and molecular evidence. Zool Anz J Comp Zool 266:169–176

Balakrishnan R (2005) Species concepts, species boundaries and species identification: a view from the tropics. Syst Biol 54:689–693

Ball SL, Armstrong KF (2006) DNA barcodes for insect pest identification: a test case with tussock moths (Lepidoptera: Lymantriidae). Can J For Res 36:337–350

Beebe NW, Foley DH, Cooper RD, Bryan JH, Saul A (1996) DNA probes for the *Anopheles punctulatus* complex. Am J Trop Med Hyg 54:395–398

Beutel RG, Pohl H (2006) Endopterygote systematics – where do we stand and what is the goal (Hexapoda, Arthropoda)? Syst Entomol 31:202–219

Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, Abebe E (2005) Defining operational taxonomic units using DNA barcode data. Philos Trans R Soc London (Ser B) 360:1935–1943

Chen B, Butlin RK, Pedro PM, Wang XZ, Harbach RE (2006) Molecular variation, systematics and distribution of the *Anopheles fluviatilis* complex in southern Asia. Med Vet Entomol 20:33–43

Claridge MF, Dawah HA, Wilson MR (eds) (1997) Species: the units of biodiversity. Chapman and Hall, London

Cock MJW, Alston-Smith S (2017) Six new records of butterflies (Lepidoptera, Papilionoidea) from Trinidad, West Indies. Living World J Trinidad Tob Field Nat Club 2017:7–13

Cywinska A, Hunter FF, Hebert PDN (2006) Identifying Canadian mosquito species through DNA barcodes. Med Vet Entomol 20:413–424

Daravath SS, Siddaiah M, ReddyaNaik B (2015) Molecular characterization and phylogenetic analysis of culex quinquefasciatus by DNA barcoding. Adv Entomol 3(3):118–124

Ebach MC, Holdrege C (2005) DNA barcoding is no substitute for taxonomy. Nature 434:697

Eberhard WG (1985) Sexual selection and animal genitalia. Harvard University Press, Cambridge, MA

Fanello CF, Santolamazza F, Torre AD (2002) Simultaneous identification of species and molecular forms of the *Anopheles gambiae* complex by PCR-RFLP. Med Vet Entomol 16:461–464

Foley DH, Wilkerson RC, Cooper RD, Volovsek ME, Bryan JH (2007) A molecular phylogeny of *Anopheles annulipes* (Diptera: Culicidae) sensu lato: the most species-rich anopheline complex. Mol Phylogenet Evol 43:283–297

Garros C, Harbach RE, Manguin S (2005) Morphological assessment and molecular phylogenetics of the Funestus and Minimus groups of *Anopheles* (*Cellia*). J Med Entomol 42:522–536

Giangrande A (2003) Biodiversity, conservation, and the 'taxonomic impediment. Aquat Conserv Mar Freshwat Ecosyst 13:451–459

Godfray HC (2002) Challenges for taxonomy. Nature 417:17–19

Green CA, Munstermann LE, Tan SG, Panyim S, Baimai V (1992) Population genetic evidence for species A, species-B, species-C and species-D of the *Anopheles dirus* complex in Thailand and enzyme electromorphs for their identification. Med Vet Entomol 6:29–36

Grimaldi DA, Engel MS (2005) Evolution of the insects. Cambridge University Press, Cambridge

Gunay F, Alten B, Simsek F, Aldemir A, Linton YM (2015) Barcoding Turkish Culex mosquitoes to facilitate arbovirus vector incrimination studies reveals hidden diversity and new potential vectors. Acta Trop 143:112–120

Gupta K, Dhawan R, Kajla M, Kumar S, Jnanasiddhy B, Singh NK, Dixit R, Bihani A, Gupta L (2016) Molecular identification of *Aedes aegypti* mosquitoes from Pilani region of Rajasthan, India. J Vector Borne Dis 53(2):149–155

Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006) DNA barcodes distinguish species of tropical Lepidoptera. Proc Natl Acad Sci USA 103(4):968–971

Hebert PDN, Gregory TR (2005) The promise of DNA barcoding for taxonomy. Syst Biol 54 (5):852–859

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003a) Biological identifications through DNA barcodes. Proc R Soc B Biol Sci 270:313–321

Hebert PDN, Ratnasingham S, deWaard JR (2003b) Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. Proc R Soc B Biol Sci 270(Suppl):S96–S99

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc Natl Acad Sci USA 101(41):14812–14817

Janzen DH, Hajibabaei M, Burns JM, Hallwachs W, Remigio E, Hebert PDN (2005) Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. Philos Trans R Soc B Biol Sci 360:1835–1845

Kent RJ, West AJ, Norris DE (2004) Molecular differentiation of colonized human malaria vectors by 28S ribosomal DNA polymorphisms. Am J Trop Med Hyg 71:514–517

Kumar NP, Rajavel AR, Natarajan R, Jambulingam P (2007) DNA barcodes can distinguish species of Indian mosquitoes (Diptera: Culicidae). J Med Entomol 44:1–7

Mallet J, Willmott K (2003) Taxonomy: renaissance or tower of babel? Trends Ecol Evol 18:57–59

Marrelli MT, Floeter-Winter LM, Malafronte RS, Tadei WP, Lourenco-de-Oliveira R, Flores-Mendoza C, Marinotti O (2005) Amazonian malaria vector anopheline relationships interpreted from ITS2 rDNA sequences. Med Vet Entomol 19:208–218

Mayden RL (1997) A hierarchy of species concepts: the denouement in the saga of the species problem. In: Claridge MF, Dawah HA, Wilson MR (eds) Species: The units of biodiversity. Chapman and Hall, London, pp 381–424

Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. PLoS Biol 3(12):2229–2238

Moritz C, Cicero C (2004) DNA barcoding: promise and pitfalls. PLoS Biol 2(10):28–30

Murugan K, Vadivalagan C, Karthika P, Panneerselvam C, Paulpandi M, Subramaniam J, Wei H, Alsalhi MS, Devanesan S, Nicoletti M, Paramasivan R (2016) DNA barcoding and molecular evolution of mosquito vectors of medical and veterinary importance. Parasitol Res 115 (1):107–121

Porter TM, Gibson JM, Shokralla S, Baird DJ, Golding GB, Hajibabaei M (2014) Rapid and accurate taxonomic classification of insect (Class Insecta) cytochrome c oxidase subunit 1 (COI) DNA barcode sequences using a naïve Bayesian classifier. Mol Ecol Resour 14(5):929–942

Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopedia of life: an introduction to DNA barcoding. Philos Trans R Soc B Biol Sci 360:1805–1811

Shendure J, Mitra RD, Varma C, Church GM (2004) Advanced sequencing technologies: methods and goals. Nat Rev Genet 5:335–344

Shokralla S, Gibson JF, Nikbakht H, Janzen DH, Hallwachs W, Hajibabaei M (2014) Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. Mol Ecol Resour 14(5):892–901

Shokralla S, Porter TM, Gibson JF, Dobosz R, Janzen DH, Hallwachs W, Golding GB, Hajibabaei M (2015) Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. Sci Rep 5:9687

Singer G, Hajibabaei M (2009) iBarcode.org: web-based molecular biodiversity analysis. BMC Bioinform 10:S14

Singh OP, Chandra D, Nanda N, Sharma SK, Htun PT, Adak T, Subbarao SK, Dash AP (2006) On the conspecificity of *Anopheles fluviatilis* species S with *Anopheles minimus* species C. J Biosci 31:671–677

Smith MA, Fisher BL, Hebert PDN (2005) DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. Philos Trans R Soc B Biol Sci 360:1825–1834

Theron J, Cloete TE (2000) Molecular techniques for determining microbial diversity and community structure in natural environments. Crit Rev Microbiol 26:37–57

Wang G, Li C, Guo X, Xing D, Dong Y, Wang Z, Zhang Y, Liu M, Zheng Z, Zhang H, Zhu X, Wu Z, Zhao T, Pinto J (2012) Identifying the main mosquito species in China based on DNA barcoding. PLoS One 7(10):e47051

Will KW, Rubinoff D (2004) Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics 20:47–55

Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. Syst Biol 54(5):844–851

DNA Barcoding of Rays from the South China Sea

B. Akbar John, M. A. Muhamad Asrul, Wahidah Mohd Arshaad, K. C. A. Jalal, and Hassan I. Sheikh

Abstract Fishery management on elasmobranch is gaining attention in recent years due to their economic value and key ecological role played by them in their natural habitat. Many species of elasmobranch especially rays exhibit overlapping morphological similarities and hence difficult to identify to their species level. As an accurate identification is the key for developing framework for fishery management, we used molecular approach to identify ray fishes sampled from South China Sea. We used cytochrome oxidase subunit I (COI) sequencing (~652 bp) to cross-examine field identification of ray fishes. A total of ten species that belong to three families were successfully sequenced/identified from 29 PCR products. BLAST/BOLD analyses were performed, and inter- and intraspecies genetic distances were calculated. Due to overlapping morphological characters and morphocryptic nature, their accurate field identification is challenging. We also addressed problems in species-level delimitation of ray fishes due to the paucity of information in DNA databanks besides unresolved taxonomic status of available data set. We further addressed management and action plan for sustainable management of elasmobranch fishery in Malaysia.

Keywords DNA barcoding · COI gene · Ray fish · Batoids

B. Akbar John (✉)
INOCEM Research Station (IRS), Kulliyyah of Science, International Islamic University Malaysia (IIUM), Kuantan, Pahang, Malaysia

M. A. Muhamad Asrul · K. C. A. Jalal
Department of Marine Science, Kulliyyah of Science, International Islamic University Malaysia (IIUM), Kuantan, Pahang, Malaysia

W. M. Arshaad
Department of Fisheries, SEFDEC, Terengganu, Malaysia

H. I. Sheikh
Department of Biotechnology, Kulliyyah of Science, International Islamic University Malaysia (IIUM), Kuantan, Pahang, Malaysia

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_14

1 Introduction

In the last century, the number of species extinction from their natural habitat has increased in alarming rate due to the effects of global climate change, habitat destruction and overharvesting. This is due to an increase in yearly rate of extinction from one species/million to ≥1000/million from the estimated 2 million identified plant and animal species, while the total estimate is ranging from 5 to 50 million animal taxa besides 10 million microorganisms (Hebert et al. 2003). Many species are extinct before they are being identified from their natural habitat. This has directed researchers and policy makers to initiate an attempt to preserve species diversity in the face of accelerating habitat destruction which have led to the awareness on the importance of accurate species identification.

In aquatic habitat, many closely related species share similar and overlapping morphological characters which could not be differentiated without the assistance from expert taxonomists. For instance, many researchers argued that elasmobranchs exhibit cryptic morphological overlapping characters, species complexity besides ontogenetic colour pattern or variation which are shared among the closely related taxa that makes the accurate identification challenging (Cerutti-Pereyra et al. 2012; Sandoval-Castillo and Rocha-Olivares 2011; Toffoli et al. 2008). It also leads to taxonomic misidentification in the field sampling and thereby effecting conservation related research and fishery management. Hence, DNA-based species identification was proposed and well established on many cryptic taxonomic forms such as lepidopteron, fishes, spiders, polychaete and many other farms (Blagoev et al. 2016; Burns et al. 2007; Hebert et al. 2003; Hebert and Gregory 2005; Steinke et al. 2016; Ward et al. 2009). Efficiency of DNA barcoding technique is also validated on ray fish identification from Taiwan (Chang et al. 2017), Indonesia (Madduppa et al. 2016), Australia (Ward et al. 2008), India (Bineesh et al. 2016) and other places (Cerutti-Pereyra et al. 2012; Toffoli et al. 2008). Though Malaysian water is well diversified with 84 species of batoids (second largest ray fish diversity in Southeast Asia), their accurate identification using DNA barcoding technique is still limited, while their field identification is highly challenging due to lacking of conventional taxonomist and problems addressed above.

Reports suggested that at least 62.5% of ecology based papers published in reputed peer journals do not provide supporting documents to justify accurate identification of organism under study (Bortolus 2008). It is also argued that, though checklist on detailed morphological characters of the species under study is available, their field identification still requires expert taxonomist to differentiate species that share overlapping morphological characters, sexual dimorphisms and ontogenetic colour pattern. Interestingly, ray fishes exhibit morphocryptic characters, species complexity and differential ontogenetic variations, and hence, in this study, we addressed the efficiency of universal DNA barcoding technique in species identification and field validation of ray fishes.

2 Materials and Methods

2.1 Sample Collection and Preparation

Ray fishes were collected from three different locations from Malaysia waters (Fig. 1). A total of 6 samples from Mukah (West Malaysia), 9 samples from Dungun (Peninsular Malaysia) and 18 samples from Kuantan (Peninsular Malaysia) were collected either by gill nets, hooks and lines. Samples were also confiscated from local market and identified morphologically using standard references (Ahmad et al. 2013; Ambak and Terengganu 2012). Tissue samples of 2 × 2 cm were excised from ventral side of each fishes and transferred into 70% ethanol then at −20 °C prior to analysis. Voucher specimens were photographed for further analysis and reference. Field identification of samples and corresponding geographical location are shown in Table 1.

2.2 DNA Barcoding and Sequencing

Total genomic DNA was extracted using DNeasy Blood and Tissue Kit (Qiagen Inc., Germany) following manufacturer guidelines. Concentration and purity of genomic DNA were analysed using NanoDrop 2000 UV-Vis Spectrophotometer. PCR amplification of target gene (652 base pairs fragments of partial cytochrome oxidase C subunit 1 gene) was carried out using multiple universal primer sets (Table 2).

Each 25 µl of PCR reaction mixture contained 9.5 µl of sterile distilled water, 12.5 µl of 1×MyTaq™ Mix (BIOLINE), 1.0 µl of each forward and reverse primer (5.0 µM) and 1.0 µl of DNA template (approximately 30–100 ng of DNA). The PCR thermal cycle conditions were set up with an initial of 2 min for denaturation at 94.0 °C, followed by 35 cycles of 30 s for denaturation at 94.0 °C, 40 sec for amplification at 52.5 °C, 1 min for annealing at 72.0 °C and finished with final extension for 10 min at 72.0 °C then held at 4.0 °C. PCR products were gel eluded on 1.5% agarose gel and visualised under AlphaImager HP system (Alpha Innotech, USA). To ensure ample amount of PCR product recovery, optimisation step was performed (if necessary) to ensure quality PCR products were used to derive DNA sequencing. Sequencing was done by outsourcing at Repfon Glamor Sdn. Bhd following Sanger sequencing standard protocol.

2.3 Data Analysis

Chromatograms were inspected for ambiguous, noisy base pairs prior to stop codon analysis. Noisy tails of sequences were trimmed, and either one of the strands was

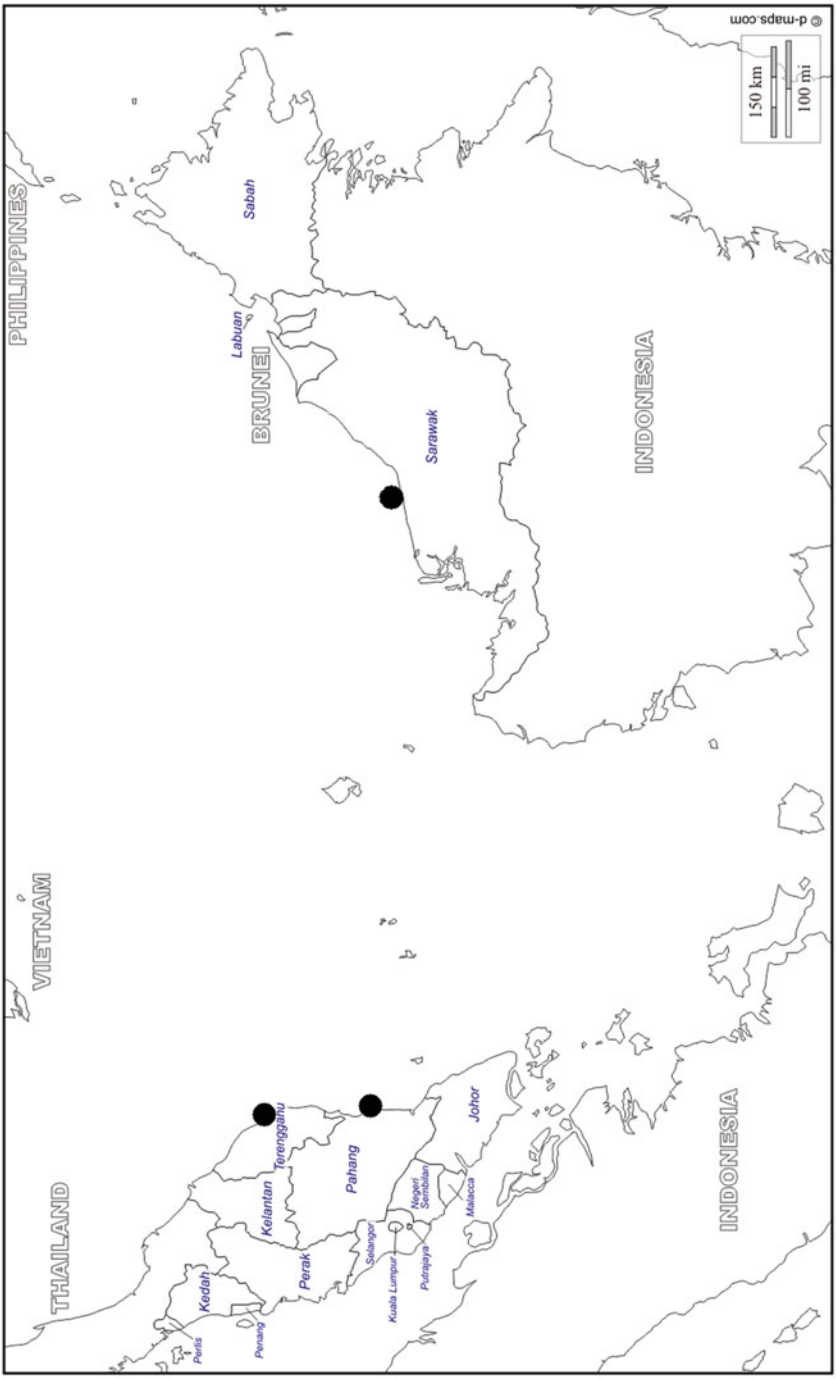


Fig. 1 Location of sampling sites

Table 1 Ray fish samples collected from three sampling locations in South China Sea

No.	Sample code	Field identification	Sampling location
1	PMKH 1	*Gymnura japonica*	Mukah, Sarawak, West Malaysia (WM) (2.5958781°N, 112.1824272°E)
2	PMKH 2	*Gymnura poecilura*	
3	PMKH 3	*Gymnura japonica*	
4	PMKH 6	*Himantura uarnacoides*	
5	PMKH 7	*Himantura uarnacoides*	
6	PMKH 8	*Himantura* cf. *uarnacoides*	
7	PDGN 1	*Gymnura poecilura**	Dungun, Terengganu, Peninsular Malaysia (PM) (4.9427071°N, 103.1772143°E)
8	PDGN 2	*Gymnura poecilura**	
9	PDGN 4	*Himantura uarnak**	
10	PDGN 5	*Himantura uarnak**	
11	PDGN 6	*Aetobatus narinari**	
12	PDGN 7	*Taeniura lymma*	
13	PDGN 8	*Taeniura lymma*	
14	PDGN 10	*Dasyatis parvonigra*	
15	PDGN 11	*Dasyatis parvonigra*	
16	PKTN 7	*Gymnura japonica*	Kuantan, Pahang, Peninsular Malaysia (PM) (3.9800417°N, 103.4098969°E)
17	PKTN 8	*Gymnura japonica*	
18	PKTN 9	*Gymnura japonica*	
19	PKTN 18	*Himantura uarnak*	
20	PKTN 22	*Aetobatus ocellatus*	
21	PKTN 28	*Himantura walga*	
22	PKTN 29	*Himantura walga*	
23	PKTN 30	*Himantura walga*	
24	PKTN 31	*Dasyatis parvonigra*	

(continued)

Table 1 (continued)

No.	Sample code	Field identification	Sampling location
25	PKTN 33	*Gymnura poecilura*	
26	PKTN 34	*Gymnura poecilura*	
27	PKTN 35	*Himantura leopard**	
28	PKTN 41	*Taeniura lymma*	
29	PKTN 52	*Taeniura lymma*	
30	PKTN 62	*Dasyatis parvonigra*	
31	PKTN 87	*Gymnura poecilura*	
32	PKTN 96	*Dasyatis parvonigra*	
33	PKTN 99	*Himantura uarnak*	

Note: (*) in scientific names represents unsure field identification by conventional taxonomist due to cryptic morphological characters. PMKH, PDGN and PKTN represent species code used in this paper for samples collected from sampling stations Mukah, Dungun and Kuantan, respectively

reverse complimented and for identifying primer targeted sequence region. Sequences were run in BLAST to identify possible species match, and the data was cross-examined with field identification of samples. Sequences with less than 500 bp were excluded for phylogenetic tree construction. Software was used for Chromas Lite (Technelysium Pty. Ltd, Australia) 2.1v for sequence editing, ClustalX (University College Dublin, Ireland) 2.1v was used for sequence alignment, and Mega6: Molecular Evolutionary Genetics Analysis Version 6.0 software was used for phylogenetic analysis. *Epinephelus lanceolatus* (GenBank accession ID: HQ174835.1) was used as an out-group to construct NJ phylogenetic tree.

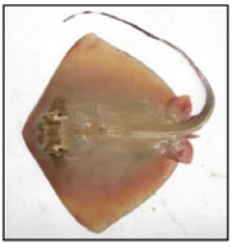
3 Results and Discussion

A total of 29 out of 33 samples were successfully barcoded for a fragment of the COI gene (~652 bp). There were ten species which belong to three families, namely, Myliobatidae (*Aetobatus ocellatus*), Gymnuridae (*Gymnura japonica*, *G. poecilura*) and Dasyatidae (*Dasyatis parvonigra*, *Neotrygon kuhlii*, *Taeniura lymma*, *Himantura leoparda*, *H. uarnacoides*, *H. uarnak and H. walga*) were successfully identified either morphologically or at molecular level (Table 3). Only 16 out of 29 samples showed best match (similarity percentage) in GenBank BLAST analysis due to insufficient

Table 2 PCR primer sets used to amplify target gene (mitochondrial COX1 gene)

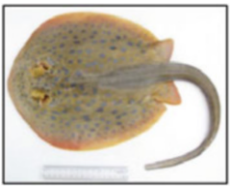
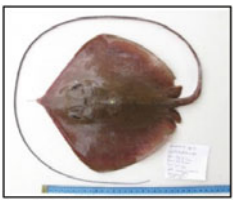
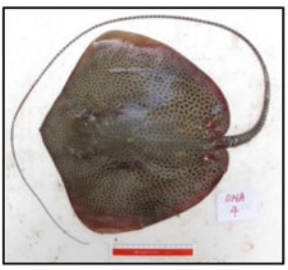
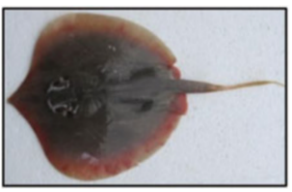
Name	Primer sequence 5′ to 3′	References
HCO2198	TAAACTTCAGGGTGACCAAAAAATCA	Folmer et al. (1994)
VF1_t1	TGTAAAACGACGGCCAGTTCTCAACCAACCACAAAGACATTGG	Ivanova et al. (2007)
VR1_t1	CAGGAAACAGCTATGACTAGACTTCTGGGTGGCCAAAGAATCA	Ivanova et al. (2007)
VF2_t1	TGTAAAACGACGGCCAGTCAACCAACCACAAAGACATTGGCAC	Ivanova et al. (2007)
FishF2_t1	TGTAAAACGACGGCCAGTCGACTAATCATAAAGATATCGGCAC	Ivanova et al. (2007)
FishR2_t1	CAGGAAACAGCTATGACACTTCAGGGTGACCGAAGAATCAGAA	Ivanova et al. (2007)
FR1d_t1	CAGGAAACAGCTATGACACCTCAGGGTGTCCGAARAAYCARAA	Ivanova et al. (2007)

Table 3 Taxonomical position and IUCN status of each species of rays sampled and barcoded in this study

Picture	Taxonomy
	Order: Myliobatiformes Family: Myliobatidae Genus: *Aetobatus* Species: *ocellatus* Common name: Eagle ray IUCN status: vulnerable Current population trend: decreasing
	Order: Myliobatiformes Family: Gymnuridae Genus: *Gymnura* Species: *japonica* Common name: Japanese Butterfly Ray IUCN status: data deficient Current population trend: unknown
	Order: Myliobatiformes Family: Gymnuridae Genus: *Gymnura* Species: *poecilura* Common name: Long tail Butterfly Ray IUCN status: near threatened Current population trend: decreasing
	Order: Myliobatiformes Family: Dasyatidae Genus: *Dasyatis* Species: *parvonigra* Common name: Dwarf Black Stingray IUCN status: data deficient Current population trend: unknown
	Order: Myliobatiformes Family: Dasyatidae Genus: *Neotrygon* Species: *kuhlii* Common name: Blue-spotted Stingray IUCN status: data deficient Current population trend: unknown

(continued)

Table 3 (continued)

Picture	Taxonomy
	Order: Myliobatiformes Family: Dasyatidae Genus: *Taeniura* Species: *lymma* Common name: Ribbontailed Stingray IUCN status: near threatened Current population trend: Unknown
	Order: Myliobatiformes Family: Dasyatidae Genus: *Himantura* Species: *leoparda* Common name: Leopard Whipray IUCN status: vulnerable Current population trend: decreasing
	Order: Myliobatiformes Family: Dasyatidae Genus: *Himantura* Species: *uarnacoides* Common name: Bleeker's Whipray IUCN status: vulnerable Current population trend: decreasing
	Order: Myliobatiformes Family: Dasyatidae Genus: *Himantura* Species: *uarnak* Common name: Reticulate Whipray IUCN status: vulnerable Current population trend: decreasing
	Order: Myliobatiformes Family: Dasyatidae Genus: *Himantura* Species: *walga* Common name: Dwarf Whipray IUCN status: near threatened Current population trend: decreasing

Note: Picture source from SEAFDEC/MFRDMD, Kuala Terengganu. Malaysia; except for *Neotrygon kuhlii* (Photographed by: Lesley Clements; Source http://www.inaturalist.org/observations/2688114)

data available in database. Hence, BOLD data analysis was used to identify remaining 13 samples and to revalidate BLAST analysed results (Tables 4 and 5). Almost all samples showed high similarity percentage (>99%) with corresponding species data available in database. However, few samples were misidentified in database similarity analysis. For instance, morphologically identified *Gymnura poecilura* was paired with *G. japonica* in BLAST analysis. Similarly *Himantura uarnak* paired with *H. leoparda* and *Aetobatus narinari* showed 100% BLAST similarity with *A. ocellatus*, while *Dasyatis parvonigra was* misidentified as *Neotrygon kuhlii*. This is due to overlapping morphological characters shared by species of same genera. For instance, *Gymnura poecilura* in Indonesian waters is identical to *G. zonura.* This has been tentatively supported by morphometric characters comparisons. In the Philippines *G. poecilura* is misidentified as *G. japonica.* Hence, the taxonomic review of Gymnuridae in the Indo-West Pacific would be helpful to resolve this problem ("The IUCN Red List of Threatened Species" 2016). Similarly *A. ocellatus* is usually misidentified as *A. narinari*. White et al. (2010) argued that these two species collected from Western Atlantic region are morphologically similar with minor morphometric differences (tail length, tail tip, background coloration of the dorsal surface) though in some samples they are overlapped. The coloration pattern generally used in field identification is highly controversial due to the direct influence of ambient environmental (water/sediment) conditions which have significant influence on skin coloration (Manjaji-Matsumoto and Last 2008). It is also decided by the age, sex, life stages, external stress and behavioural adaptability of fishes (Sandoval-Castillo and Rocha-Olivares 2011; Toffoli et al. 2008; Wynen et al. 2009). Neighbour Joining (NJ) tree showed *Dasyatis parvonigra* to be the sister taxa to *Neotrygon kuhlii*. However, morphological features can clearly distinguish the later from former as *D. parvonigra* has no spot on the upper disc in contrast with *N. kuhlii* that has blue spots. In the case of *Himantura* sp., many researchers have argued that an existence of species complex and considerable morphological variability within the same forms lead to misidentification in the field and hence urgently require taxonomic revision especially on *H. gerrardi* (Manjaji-Matsumoto and Last 2008; Ward et al. 2008, 2009).

Field identification of ray fish by expert was in line with DNA barcoding techniques to some extant due to overlapping morphological characters among the cryptic species besides limitations on the data availability in DNA databases. However, efficiency of universal barcode gene in species delimitation was apparent in this study. Similarly, previous studies have proposed mitochondrial COI gene and NADH2 gene as reliable molecular markers for identification of elasmobranches (Holmes et al. 2009; Naylor et al. 2012). Efficiency of this method in illegal species trading and confiscation besides consumer preference towards ray products sold in markets are well established (Griffiths et al. 2013). DNA barcoding is also possible when only part of an organism is available (John et al. 2016). In such cases, the method is useful in animal forensics (Holmes et al. 2009). In our study, not all field identified samples were accurately matched with same species data available in databanks, if Zemlak et al. (2009) rule of thumb for discriminating species (similarity below 96.5%) used. Thus, expert taxonomical identification of samples in field would help in ensuring accuracy in species discrimination. It should also be noted

Table 4 BLAST analysis of generated DNA sequences of ray fishes shows similarity percentage and accurate identification of species (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch)

Sample code	Field identification	Identity ratio (%)	Sequence ID (bp)	Species identified
PMKH 1	*Gymnura japonica*	652/652 (100%)	EU398804.1 (652)	**Gymnura poecilura*
PMKH 2	*Gymnura poecilura*	652/652 (100%)	EU398804.1 (652)	*Gymnura poecilura*
PMKH 3	*Gymnura japonica*	652/652 (100%)	EU398804.1 (652)	**Gymnura poecilura*
PMKH 6	*Himantura uarnacoides*	651/652 (99%)	KM073009.1 (652)	*Himantura uarnacoides*
PMKH 7	*Himantura uarnacoides*	652/652 (100%)	KM073009.1 (652)	*Himantura uarnacoides*
PMKH 8	*Himantura* cf. *uarnacoides*	651/652 (98%)	KM073009.1 (652)	**Himantura uarnacoides*
PDGN 1	*Gymnura poecilura**	652/652 (100%)	EU398804.1 (652)	*Gymnura poecilura*
PDGN 2	*Gymnura poecilura**	651/652 (99%)	EU398804.1 (652)	*Gymnura poecilura*
PDGN 4	*Himantura uarnak**	651/652 (99%)	KM072997.1 (652)	**Himantura leoparda*
PDGN 5	*Himantura uarnak**	650/652 (99%)	KM072997.1 (652)	**Himantura leopard*
PDGN 6	*Aetobatus narinari**	652/652 (100%)	EU398508.1 (652)	**Aetobatus ocellatus*
PDGN 7	*Taeniura lymma*	652/652 (100%)	KM073027.1 (652)	*Taeniura lymma*
PDGN 8	*Taeniura lymma*	652/652 (100%)	KM073027.1 (652)	*Taeniura lymma*
PDGN 11	*Dasyatis parvonigra*	652/652 (100%)	EU398733.1 (655)	**Neotrygon kuhlii*
PKTN 7	*Gymnura japonica*	651/652 (99%)	EU398804.1 (652)	**Gymnura poecilura*
PKTN 8	*Gymnura japonica*	649/652 (99%)	EU398804.1 (652)	**Gymnura poecilura*
PKTN 9	*Gymnura japonica*	652/652 (100%)	EU398804.1 (652)	**Gymnura poecilura*
PKTN 18	*Himantura uarnak*	651/652 (99%)	KM072997.1 (652)	**Himantura leoparda*
PKTN 22	*Aetobatus ocellatus*	648/652 (99%)	EU398508.1 (652)	*Aetobatus ocellatus*
PKTN 28	*Himantura walga*	651/652 (99%)	KM072995.1 (652)	*Himantura walga*
PKTN 29	*Himantura walga*	651/652 (99%)	KM072995.1 (652)	*Himantura walga*
PKTN 30	*Himantura walga*	652/652 (100%)	EU398873.1 (655)	*Himantura walga*

(continued)

Table 4 (continued)

Sample code	Field identification	Identity ratio (%)	Sequence ID (bp)	Species identified
PKTN 31	*Dasyatis parvonigra*	652/652 (100%)	EU398733.1 (655)	**Neotrygon kuhlii*
PKTN 33	*Gymnura poecilura*	652/652 (100%)	EU398804.1 (652)	*Gymnura poecilura*
PKTN 34	*Gymnura poecilura*	649/652 (99%)	EU398804.1 (652)	*Gymnura poecilura*
PKTN 35	*Himantura leoparda**	651/652 (99%)	KM072997.1 (652)	*Himantura leopard*
PKTN 41	*Taeniura lymma*	652/652 (100%)	KM073027.1 (652)	*Taeniura lymma*
PKTN 52	*Taeniura lymma*	652/652 (100%)	KM073027.1 (652)	*Taeniura lymma*
PKTN 62	*Dasyatis parvonigra*	652/652 (100%)	EU398733.1 (655)	**Neotrygon kuhlii*

Note: Symbol (*) represents the species that are not consistent with the initial identification

Table 5 The reconfirmed species after submitting the data on the BOLD system (http://www.boldsystems.org/index.php/IDS_OpenIdEngine)

Sample code	Field identification	Similarity (%)	Species identified by BOLD database
PMKH 1	*Gymnura japonica*	100	**Gymnura poecilura*
PMKH 3	*Gymnura japonica*	100	**Gymnura poecilura*
PMKH 8	*Himantura* cf. *uarnacoides*	99.85	**Himantura uarnacoides*
PDGN 4	*Himantura uarnak**	99.84	**Himantura leoparda*
PDGN 5	*Himantura uarnak**	99.84	**Himantura leoparda*
PDGN 6	*Aetobatus narinari**	100	**A. narinari/ocellatus*
PDGN 11	*Dasyatis parvonigra*	100	**Neotrygon kuhlii*
PKTN 7	*Gymnura japonica*	99.85	**Gymnura poecilura*
PKTN 8	*Gymnura japonica*	99.53	**Gymnura poecilura*
PKTN 9	*Gymnura japonica*	100	**Gymnura poecilura*
PKTN 18	*Himantura uarnak*	100	**Himantura leopard*
PKTN 31	*Dasyatis parvonigra*	100	**Neotrygon kuhlii*
PKTN 62	*Dasyatis parvonigra*	100	**Neotrygon kuhlii*

that taxonomic decision based on single molecular marker (e.g. single gene sequence [COI gene] in this study) that is maternal inherited might not resolve all species identification. Hence, it is argued that COI gene sequencing cannot be used to discriminate recently evolved species as haplotypes are shared between trans-boundary species (Toffoli et al. 2008).

The paucity of information especially on DNA data set of cartilaginous fishes in major databases (GenBank, BOLD database) perhaps would be the reason for mismatching of species. Similarly Cerutti-Pereyra et al. (2012) observed only 19 out of 67 ray fish sequences tested had consistent matches on major databases. They further argued that this might be due to regular anomalies among closely related cryptic species that makes the species identification challenging. Misidentification might also be due to the presence of species complex, miss identification in the field or unavailability of species specific sequence in databases. On the other hand, limitations associated with the online databases also include (1) significant number of taxonomically unverified entries on GenBank and BOLD databases and (2) the presence of inconsistent, incorrect, short length duplicated lodged sequences with misleading scientific names.

Phylogenetic and phylogeographical signals are apparent in NJ tree constructed in this study (Fig. 2). All 33 sequences used in this study (including out group ID: HQ174835; *G. japonica, A. narinari* and *H. uarnak*) clearly segregated into their corresponding congeneric forms. The average intraspecies genetic distance was lower (0.25%) than interspecies difference (2.45%). Same species sampled from different sampling zones has segregated into their respective clad. Low genetic variation was observed in same species collected from geographically closer population (especially same species from Kuantan and Terengganu, Peninsular Malaysia) compared to geographically distant population (Sarawak, East Malaysia). Similar observation was noted in earlier study by Cerutti-Pereyra et al. (2012) who observed higher genetic distance between samples of same species complex collected from western and eastern coasts.

3.1 Management and Action Plan

As ray fishes are considered to be one of the common table fish in Malaysia, their management and action plan for evaluating their natural stock and fishery pressure are limited compared to shark fishery. Until recently, 84 species of ray fishes were recorded in Malaysian waters compared to the 106 species of rays recorded in Indonesia that makes Malaysia to be at the second position in Southeast Asia. Due to high fishery pressure and over exploitation of selected ray fishes (manta ray), their catch is banned in Indonesia since 2012. In Malaysia, landing of elasmobranch from 1982 to 2012 contributed an average of 1.8% of total demersal fish landing. In general, fishery practice in Malaysia does not target shark and ray fishing. They are being caught by trawls accidentally and brought to the shore for reasonable sales. High consideration was drawn on the management of sharks in Malaysia through national conservation management plan (PLAN 2). The definition of the word "*shark*" includes all cartilaginous or chondrichthyan fishes, comprising sharks, rays, skates and chimaeras. This management plan includes review on shark population, trading and resource overview, revision, monitoring and evaluation for sustainable fishery management.

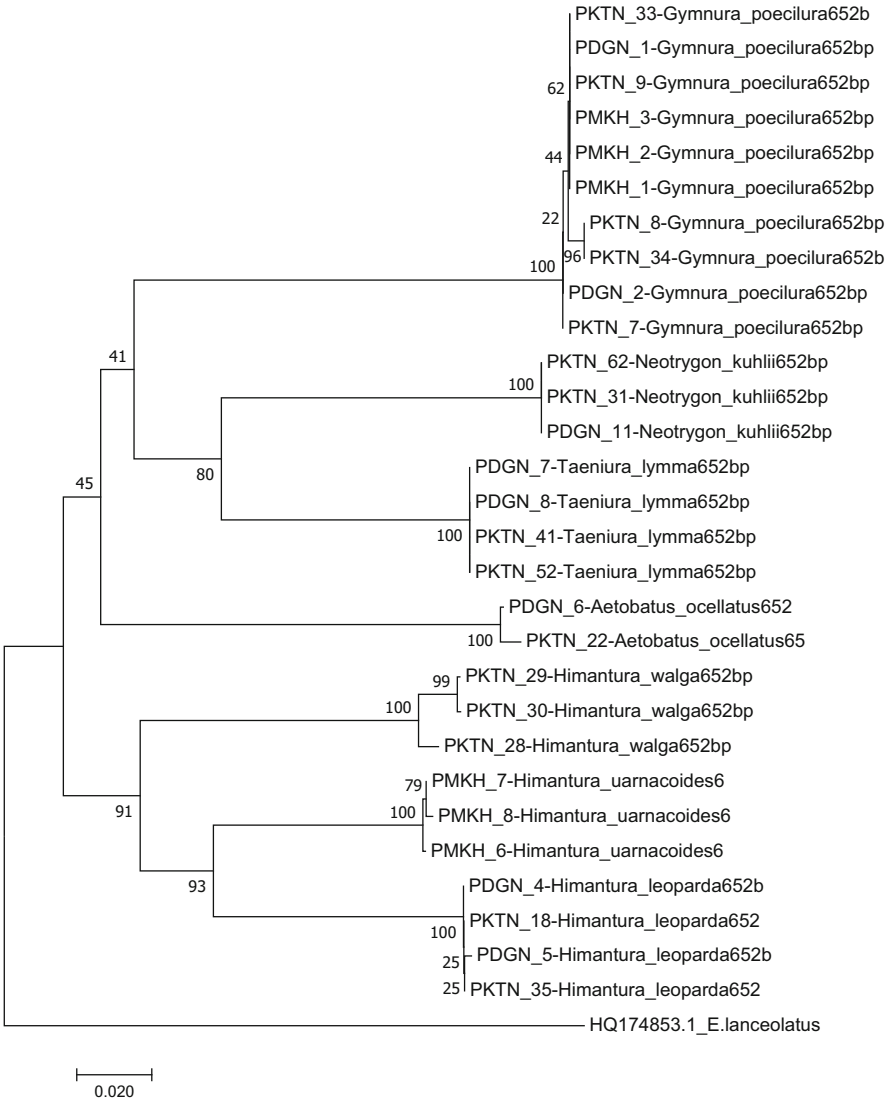


Fig. 2 Neighbour Joining (NJ) phylogenetic tree constructed to determine the evolutionary history of ray fishes. Note: The evolutionary history was inferred using the Neighbour Joining method. The optimal tree with the sum of branch length = 1.03083732 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method and are in the units of the number of amino acid differences per site. The analysis involved 34 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 620 positions in the final data set. Evolutionary analyses were conducted in MEGA6

4 Conclusion

In conclusion, DNA barcoding is an accurate and rapid technique in successful validation of consensus field identified samples. There is an urgent need of taxonomic revision on morphologically similar cryptic ray species. Misidentification is perhaps due to the availability of incongruent, inconsistent and ambiguous database in DNA databanks. Though there is a severe criticism on the efficiency of DNA barcoding techniques using single molecular marker sequencing, such approach is still highly useful in species validation and thereby can be used in various fishery management practices.

References

Ahmad A, Annie Lim PK, Fahmi D, Krajangdara T (2013) Field guide to look-alike sharks and rays species of the Southeast Asian Region. Southeast Asians Fisheries Development Center/Marine Fishery Resources Development and Management Department

Ambak MA, Terengganu, PUM (2012) Fishes of Malaysia: Penerbit UMT

Bineesh K, Gopalakrishnan A, Akhilesh K, Sajeela K, Abdussamad E, Pillai N, et al (2016) DNA barcoding reveals species composition of sharks and rays in the Indian commercial fishery. Mitochondrial DNA Part A, pp 1–15

Blagoev GA, deWaard JR, Ratnasingham S, deWaard SL, Lu L, Robertson J et al (2016) Untangling taxonomy: a DNA barcode reference library for Canadian spiders. Mol Ecol Resour 16(1):325–341

Bortolus A (2008) Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology. AMBIO 37(2):114–118

Burns JM, Janzen DH, Hajibabaei M, Hallwachs W, Hebert PD (2007) DNA barcodes of closely related (but morphologically and ecologically distinct) species of skipper butterflies (Hesperiidae) can differ by only one to three nucleotides. J Lepidopter Soc 61(3):138–153

Cerutti-Pereyra F, Meekan MG, Wei N-WV, O'Shea O, Bradshaw CJ, Austin CM (2012) Identification of rays through DNA barcoding: an application for ecologists. PLoS One 7(6):e36479

Chang CH, Shao KT, Lin HY, Chiu YC, Lee MY, Liu SH, Lin PL (2017) DNA barcodes of the native ray-finned fishes in Taiwan. Mol Ecol Resour 17:796–805

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol 3(5):294–299

Griffiths AM, Miller DD, Egan A, Fox J, Greenfield A, Mariani S (2013) DNA barcoding unveils skate (Chondrichthyes: Rajidae) species diversity in 'ray' products sold across Ireland and the UK. PeerJ 1:e129

Hebert PD, Gregory TR (2005) The promise of DNA barcoding for taxonomy. Syst Biol 54(5): 852–859

Hebert PD, Cywinska A, Ball SL (2003) Biological identifications through DNA barcodes. Proc R Soc Lond B Biol Sci 270(1512):313–321

Holmes BH, Steinke D, Ward RD (2009) Identification of shark and ray fins using DNA barcoding. Fish Res 95(2):280–288

Ivanova NV, Zemlak TS, Hanner RH, Hebert PD (2007) Universal primer cocktails for fish DNA barcoding. Mol Ecol Notes 7(4):544–548

John BA, Sheikh HI, Jalal K, Zaleha K, Kamaruzzaman B (2016) Revised phylogeny of extant xiphosurans (Horseshoe Crabs). In: DNA barcoding in marine perspectives. Springer, pp 113–130

Madduppa H, Ayuningtyas RU, Subhan B, Arafat D (2016) Exploited but unevaluated: DNA barcoding reveals skates and stingrays (Chordata, Chondrichthyes) species landed in the Indonesian fish market. Indon J Mar Sci 21(2):77–84

Manjaji-Matsumoto BM, Last PR (2008) Himantura leoparda sp. nov., a new whipray (Myliobatoidei: Dasyatidae) from the Indo–Pacific. Descriptions of new Australian Chondrichthyans. CSIRO Mar Atm Res Pap 22:293–301

Naylor GJ, Caira JN, Jensen K, Rosana K, White WT, Last P (2012) A DNA sequence-based approach to the identification of shark and ray species and its implications for global elasmobranch diversity and parasitology. Bull Am Mus Nat Hist:1–262

Sandoval-Castillo J, Rocha-Olivares A (2011) Deep mitochondrial divergence in Baja California populations of an aquilopelagic elasmobranch: the golden cownose ray. J Hered 102(3): 269–274

Steinke D, Prosser SW, Hebert PD (2016) DNA barcoding of marine metazoans. Mar Genomics Methods Protoc:155–168

The IUCN Red List of Threatened Species. *Version 2016-3.* Retrieved 28 Jan 2017 from www.iucnredlist.org

Toffoli D, Hrbek T, Araújo MLG d, Almeida MP d, Charvet-Almeida P, Farias IP (2008) A test of the utility of DNA barcoding in the radiation of the freshwater stingray genus Potamotrygon (Potamotrygonidae, Myliobatiformes). Genet Mol Biol 31(1):324–336

Ward RD, Holmes BH, White WT, Last PR (2008) DNA barcoding Australasian chondrichthyans: results and potential uses in conservation. Mar Freshw Res 59(1):57–71

Ward RD, Hanner R, Hebert PD (2009) The campaign to DNA barcode all fishes, FISH-BOL. J Fish Biol 74(2):329–356

White WT, Last PR, Naylor GJ, Jensen K, Caira JN (2010) Clarification of Aetobatus ocellatus (Kuhl, 1823) as a valid species, and a comparison with Aetobatus narinari (Euphrasen, 1790) (Rajiformes: Myliobatidae). Descriptions of new sharks and rays from Borneo. CSIRO Mar Atmos Res Pap 32:141–164

Wynen L, Larson H, Thorburn D, Peverell S, Morgan D, Field I, Gibb K (2009) Mitochondrial DNA supports the identification of two endangered river sharks (Glyphis glyphis and Glyphis garricki) across northern Australia. Mar Freshw Res 60(6):554–562

Zemlak TS, Ward RD, Connell AD, Holmes BH, Hebert PD (2009) DNA barcoding reveals overlooked marine fishes. Mol Ecol Resour 9(s1):237–242

Molecular Phylogeny of Elasmobranchs

A. Pavan-Kumar, P. Gireesh-Babu, A. K. Jaiswar, S. G. Raje, A. Chaudhari, and G. Krishna

Abstract Elasmobranchs (sharks, rays and skates) are considered as one of the most ancient and successful vertebrate lineages. According to fossil evidence, elasmobranchs were present from the lower Devonian period and diversified during the "Age of Fishes" to become widely distributed and represented by diverse morphological and ecological forms. The elasmobranch lineage has survived four mass extinction events, and the most extant taxa are derivable from Mesozoic forms. Despite a highly successful evolutionary trajectory, elasmobranch species have not been characterized completely, and still ambiguity persists for some of the species identification. Phylogenetic studies with molecular data would resolve taxonomic ambiguity and provide insights in the evolutionary relationship among elasmobranchs. This chapter summarizes the complete phylogeny work carried on elasmobranchs and highlights the significance of phylogeny for formulating conservation measures.

Keywords Mitochondrial markers · Nuclear markers · Taxonomy · Conservation

1 Introduction

The elasmobranchs comprising sharks, skates and rays are considered as one of the most ancient and successful vertebrate lineages. According to fossil evidence, elasmobranchs were present from the lower Devonian period, diversified during

A. Pavan-Kumar (✉) · P. Gireesh-Babu · A. Chaudhari · G. Krishna
Fish Genetics and Biotechnology Division, ICAR-Central Institute of Fisheries Education, Mumbai, Maharashtra, India
e-mail: pavankumar@cife.edu.in

A. K. Jaiswar
Fisheries Resources and Post Harvest Division, ICAR-Central Institute of Fisheries Education, Mumbai, Maharashtra, India

S. G. Raje
ICAR-Central Marine Fisheries Research Institute, Regional Centre, Mumbai, Maharashtra, India

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_15

the "Age of Fishes" and represented by diverse morphological and ecological forms (Grogan and Lund 2004). Due to their basal position in the vertebrate evolutionary tree, elasmobranchs have been used as model organism to study primitive vertebrate physiology, comparative biology and comparative genomics (Venkatesh et al. 2007; Marra et al. 2017). As apex predators they play an important role in ecological balance. However, globally elasmobranch stocks are declining, and several species conservation status has been evaluated as critical/endangered (IUCN 2016). Proper conservation measures are warranted to restore the depleted stocks. For effective formulation of conservation measures, it is essential to have stable taxonomy to delineate species accurately.

Classification based on morphology grouped elasmobranchs into two monophyletic groups: batoids and sharks (Regan 1906; White 1937). Later, based on phenetics (similarities), Compagno (1973a, 1977) classified elasmobranchs into four separate superorders, namely, Galeomorphii (Orders: Heterodontiformes, Carcharhiniformes, Lamniformes, Orectolobiformes), Squalomorphii (Orders: Hexanchiformes, Pristiophoriformes, Squaliformes), Squatinomorphii (Genus *Squatina*) and Batoidea (Orders: Rajiformes, Rhinobatiformes, Myliobatiformes, Torpediniformes, Pristiformes) (Fig. 1a, b). Maisey (1980) merged superorders squalomorphs and squatinomorphs as orbitostylic sharks on the basis of potential synapomorphic trait (an orbital process that projects from the upper jaw cartilage inside the eye socket) (Fig. 2).

Later several phylogenetic studies based on morphological characteristics (external, skeletal and muscular) reported that batoids (rays and skates) were derived from sharks and grouped them with the Pristiophoriformes and Squatinomorphii. This clade (Hypnosqualea) along with other squalomorphs has been termed as "Squalea" and formed as a sister group to galeomorphs (Fig. 3) (Shirai 1992, 1996; de Carvalho 1996; de carvalho and Maisey 1996). The synapomorphic traits considered for this classification were orbital articulation, a basal angle on the suboptimal cranium and widely separated nasal capsules.

Due to some inherent problems with morphological characteristics such as poor preservation of fossil endoskeleton, convergent evolution of morphological traits and paucity of recognizable synapomorphies for certain groups, molecular markers (nuclear and mitochondrial) have been implied to infer phylogenetic relationship among orders/families of elasmobranchs (Bernardi et al. 1992; Douady et al. 2003; Dosay-Akbulut 2008).

2 Interordinal Phylogeny

Several phylogenetic studies based on mitochondrial and nuclear genes supported major division of sharks into Squalimorphii and Galeomorphii (Douady et al. 2003; Heinicke et al. 2009). Most of the molecular phylogenetic studies rejected hypnosqualean hypothesis that positioned batoids within sharks (Compagno 1977; de Carvalho and Maisey 1996; Shirai 1992, 1996). Slight variation was observed in

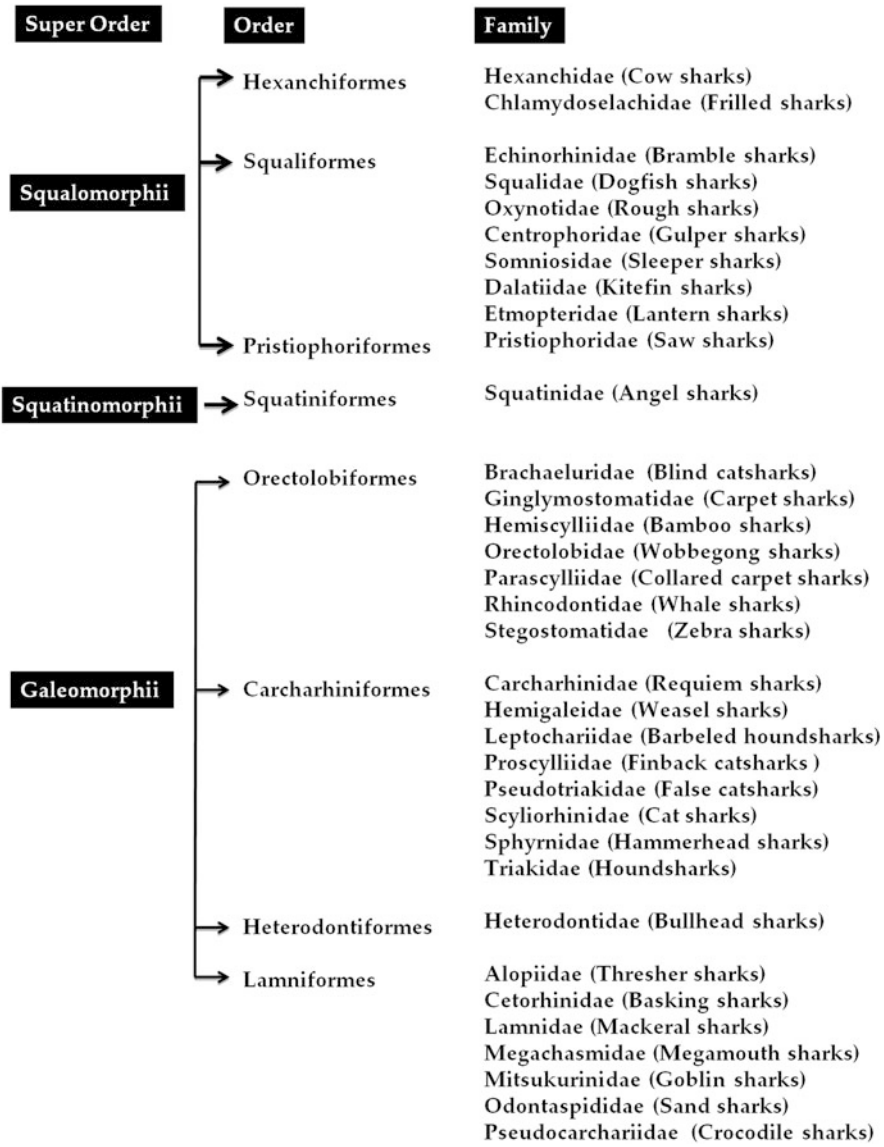


Fig. 1 (**a**) Classification of sharks, (**b**) classification of rays (Batoidea) (asterisk, the place of this family is to be confirmed)

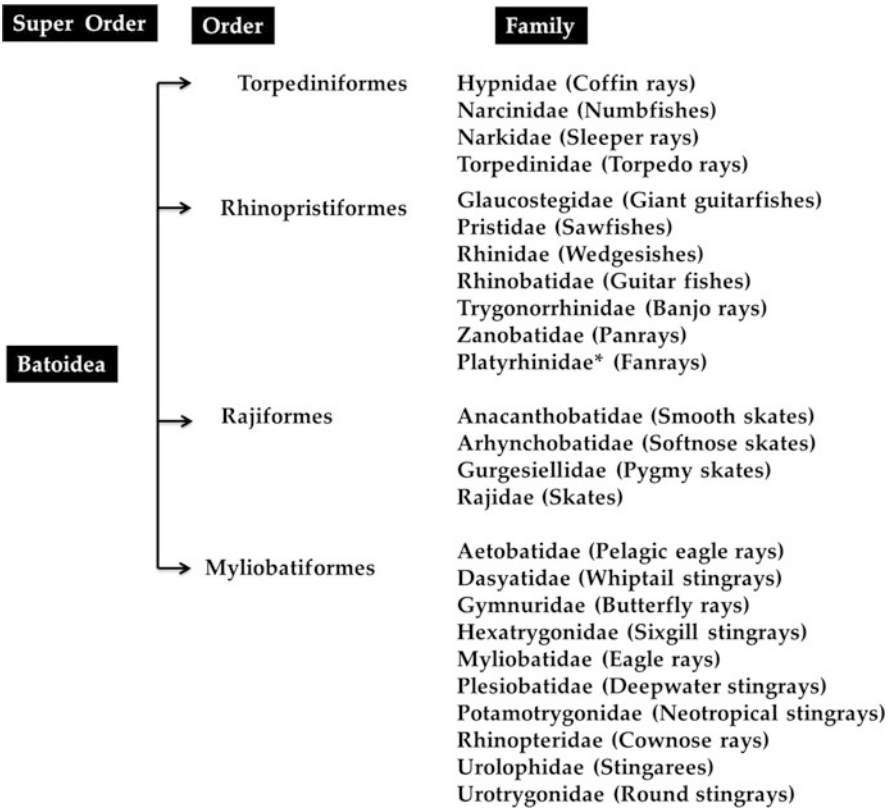


Fig. 1 (continued)

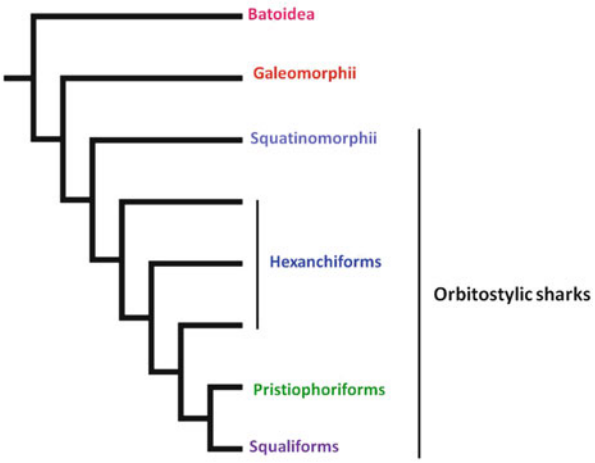


Fig. 2 Hypothesis of elasmobranch relationships proposed by Maisey (1980)

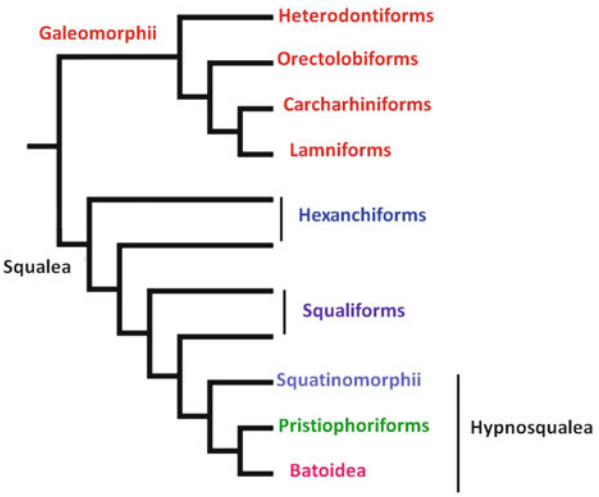


Fig. 3 Hypothesis of elasmobranch relationships proposed by Shirai (1992, 1996)

tree topologies generated by different data sets such as mitochondrial genes or concatenated mitochondrial and nuclear genes (Naylor et al. 2005, 2012; Pavan-Kumar et al. 2014; Vélez-Zuazo and Agnarsson 2011). Based on concatenated data, monophyly was observed for Hexanchiformes and Squatiniformes whereas Pristiophoriformes and Squaliformes are non-monophyletic. Phylogeny tree showed sister relation of Hexanchiformes to other orders. The relationship within Squalimorphii was defined as {Hexanchiformes [Squaliformes (Pristiophoriformes, Squatiniformes)]}. Within Galeomorphs, Carcharhiniformes and Orectolobiformes formed as sister group. Orders Lamniformes, Heterodontiformes and Carcharhiniformes are monophyletic, while Orectolobiformes is paraphyletic (Vélez-Zuazo and Agnarsson 2011). However, phylogeny tree constructed by mitochondrial genes proposed different relationship among these orders (Naylor et al. 2012; Pavan-Kumar et al. 2014). Within Galeomorphii, sister group relation was observed between Lamniformes and Orectolobiformes, and this clade is joined by Carcharhiniformes and then Heterodontiformes (Fig. 4) (Winchell et al. 2004; Pavan-Kumar et al. 2014).

3 Batoidea

Batoids are characterized by dorsoventrally flattened body, greatly expanded pectoral fins attached to the side of the head and ventral mouth. Batoids include Torpediniformes (electric rays), Pristiformes (sawfishes), Rhinobatiformes (guitarfishes), Rajiformes (skates) and Myliobatiformes (stingrays) (McEachran and Aschliman 2004; Nelson 2006). Several hypotheses based on morphological characters have been proposed, and no consensus agreement was observed among all these hypotheses (Compagno

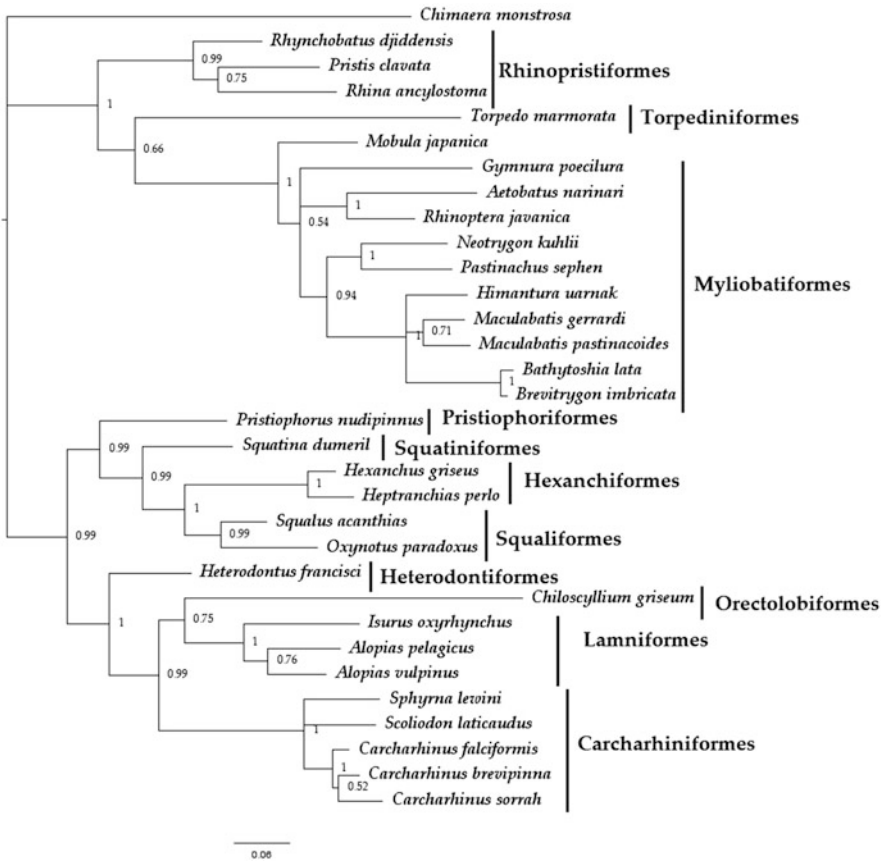


Fig. 4 Bayesian inference based phylogeny depicting relationship among major orders of elasmobranch inferred from mitochondrial markers (adapted from Pavan-Kumar et al. 2014)

1973b, 1977, 1999; Heemstra and Smith 1980; Nishida 1990; Shirai 1992, 1996; McEachran et al. 1996; McEachran and Dunn 1998; Peach and Rouse 2004). Cladistic analysis based on morphological characters supported the derivation of batoids from sharks and placed them within sharks (de Carvalho 1996; Shirai 1996). However, molecular phylogenetic analysis disproved this hypothesis and proposed batoids as separate group from sharks (Lawson et al. 1995; Douady et al. 2003; Winchell et al. 2004).

Molecular phylogenetic analyses revealed monophyletic nature of all orders except guitarfishes (Naylor et al. 2012). Skates (Rajiformes) were recovered as sister to all other batoids, and sawfish (*Pristis clavata*) was nested within the guitarfish clades. Whereas morphological studies suggested that sawfishes are sister to all other batoids (Nishida 1990; Shirai 1992, 1996; Kriwet 2004). Based on phylogenetic analysis, families Pristidae (sawfishes), Rhinidae (sharkrays), Rhynchobatidae (wedgefish), Rhinobatidae (guitarfish) and Zanobatidae were grouped under an order Rhinopristiformes (Naylor et al. 2012).

Evolutionary relationships among the four major lineages of batoids (Rajiformes, Torpediniformes, Rhinopristiformes and Myliobatiformes) have shown slight variations as per number of taxa and genes used for reconstructing phylogeny (Naylor et al. 2012). In few studies, Torpediniformes formed as a basal to a group consisting of the Myliobatiformes and the Rhinopristiformes + Rajiformes (McEachran et al. 1996; Rocco et al. 2007). Shirai (1996) proposed the Rhinopristiformes to be basal to a group consisting of the Torpediniformes and the Rajiformes + Myliobatiformes.

4 Intraordinal Phylogeny

4.1 Carcharhiniformes

Carcharhiniformes (ground sharks) is the largest order of sharks with more than 270 species classified into eight families. The monophyletic nature of this order has been confirmed and supported by morphological characters as well as molecular data (Douady et al. 2003; Winchell et al. 2004; Iglésias et al. 2005). Phylogeny studies using mitochondrial partial 12S ribosomal RNA, valine tRNA, 16S rRNA and RAG gene revealed that families Carcharhinidae, Proscyllidae, Pseudotriakidae and Triakidae are sister group to the scyliorhinids subgroup that consist of species of *Apristurus*, *Asymbolus*, *Cephalurus*, *Galeus* and *Parmaturus*. Whereas other subgroup of scyliorhinids comprising species of *Scyliorhinus* and *Cephaloscyllium* formed as a basal branch within Carcharhiniformes (Iglésias et al. 2005). This study has highlighted paraphyly within scyliorhinids and triakids.

4.2 Squaliformes

Squaliform sharks are highly specialized sharks with more than 130 species classified into 24 genera. Some of the squaliform species have been reported to have bioluminescence, and this character plays a vital role in evolution and speciation of these sharks. Straube et al. (2015) investigated evolutionary relationship among squaliform taxa using single-copy nuclear exons. The phylogenetic analysis suggested that family Echniorhinidae does not come under squaliform sharks but are sister group to Squatiniformes and Pristiophoriformes. In phylogenetic tree, Squalidae formed as basal group whereas Etmopteridae was the derived group. Oxynotidae nested within Somniosidae indicating paraphyletic nature of the later family. Family Centrophoridae split from Etmopteridae, Dalatiidae, Somniosidae and Oxynotidae. Dalatiidae were sister group to clade comprising Etmopteridae, Somniosidae and Oxynotidae.

4.3 *Myliobatiformes*

Myliobatiformes (stingrays) includes around 170 species and has wide distribution from tropical to temperate waters (Nelson 2006). Several morphological synapomorphies supported monophyly of Myliobatiformes. These ray fishes were grouped into four major clades, namely, Urolophidae, Dasyatidae, Gymnuridae and Myliobatidae (Nishida 1990). Later, McEachran et al. (1996) classified Myliobatiformes into three major clades (Urotrygonidae, Potamotrygonidae and Dasyatidae) with a basal taxon of *Hexatrygon*, *Plesiobatis* and *Urolophus*. Phylogenetic analyses using mitochondrial 12S and 16S rRNA genes showed tree where *Urolophus* formed as basal group within myliobatids. *Myliobatis* and *Urobatis* formed as sister group to the clades of Rhinoptera + Mobulidae and Potamotrygon + Himantura, respectively (Lovejoy 1996; Dunn et al. 2003; McEachran et al. 1996).

5 Family-Level Phylogeny

5.1 *Orectolobidae*

Wobbegong sharks are dorsoventrally flattened demersal sharks with unique dermal lobes on the lateral sides of the head (Compagno 2001). Orectolobidae comprises three genera, namely, *Eucrossorhinus* Regan, 1908 (*Eucrossorhinus dasypogon*), *Sutorectus* Whitely, 1939 (*Sutorectus tentaculatus*) and *Orectolobus* Bonaparte, 1834 (*O. japonicus*, *O. wardi*, *O. maculates*, *O. ornatus*, *O. halei*, *O. hatchinsi*, *O. floridus*, *O. parvimaculatus*). Several morphology-based studies placed *Sutorectus* in ancestral clade and showed that *Eucrossorhinus* was the most derived genus and sister to *Orectolobus* (Dingerkus 1986; Goto 2001). Molecular phylogeny analyses using four mitochondrial and nuclear genes revealed *Eucrossorhinus* at the basal position and *Sutorectus* as a sister group to *O. floridus* in a recently derived clade (Corrigan and Beheregaray 2009). Based on the genetic divergence values and evolutionary relationship, this study has recommended to revise the taxonomy of this family to include all species within in a single monophyletic genus, *Orectolobus*.

5.2 *Scyliorhinidae*

Sharks of the family Scyliorhinidae (Order: Carcharhiniformes) are known as catsharks and consist of approximately 17 genera with 151 species (Compagno 2005). These species were classified into two subfamilies: Scyliorhininae and Pentanchinae (tribes Pentanchini and Halaelurini). Phylogeny analyses using cytochrome *b* (cyt *b*), nicotinamide adenine dehydrogenase subunit 2 (NADH-2) and NADH4 genes revealed paraphyletic nature of Scyliorhinidae family. Clade of Pentanchini and Halaelurini

showed sister group relation to the order Carcharhiniformes. This study supported Halaelurini hypotheses and showed sister relationship between *Haploblepharus pictus* and *H. fuscus* while Halaelurus formed as basal to Haploblepharus (Human et al. 2006).

5.3 Sphyrnidae

Due to the laterally expanded, dorsoventrally compressed head (cephalofoil), species of family Sphyrnidae are known as hammerhead sharks. The structure and shape of cephalofoil is different in various groups of sharks (Mara et al. 2015). Laterally minimum expanded head is observed in the bonnethead shark (*Sphyrna tiburo*), while the winghead shark (*Eusphyra blochii*) has the most laterally expanded head. Based on the body size and head shape, Gilbert (1967) has classified these sharks into three genera, namely, *Sphyrna* (large body size), *Platysqualus* (small body size) and *Eusphyra* (head shape). Later Compagno (1988) modified this classification by erecting a new subgenus *Mesozygaena* to accommodate small body size species (*S. corona*, *S. media* and *S. tudes*). He placed *Platysqualus* as a subgenus of *Sphyrna* and included *S. tiburo* under this subgenus. All large species were placed under subgenus *Sphyrna*. In Compagno's phylogenetic tree, *S. tiburo* (narrowest head) was placed as a basal group, and *E. blochii* (laterally expanded head) was deeply nested within tree indicating the extant species represents elaboration of head shape (Compagno 1988). Lim et al. (2010) used sequences of nuclear internal transcribed spacer (ITS2), distal-less 1 and 2 (Dlx1 and Dlx2) and mitochondrial ND2, cytochrome *b*, cytochrome *c* oxidase I and D-loop for reconstructing evolutionary history among species of Sphyrnidae family. The resultant phylogeny tree showed divergent evolution of cephalofoil among different lineages. This study also proved that both the small and large hammerhead sharks are not monophyletic. This study supported hypothesis that the common ancestor of all extant hammerhead sharks was large (>150 cm), and subsequently the cephalofoil might have evolved as a hydrodynamic adaptation (Lim et al. 2010).

5.4 Triakidae

Species of family Triakidae (houndsharks) are classified into two subfamilies, viz. Triakinae (Genera: *Triakis*, *Mustelus*, *Scylliogaleus*) and Galeorhininae. The species of Galeorhininae are further divided into iagins (*Iago*, *Furgaleus*, *Hemitriakis*, and *Gogolia*) and Galeorhininis (*Galeorhinus* and *Hypogaleus*). Lopez et al. (2006) analysed phylogenetic relationship using the mitochondrial genes cytochrome *b* (Cyt *b*) and NADH subunit 2 (NADH2) and NADH4 and the single-copy nuclear gene recombination-activating gene 1 (RAG1). This study results showed sister group relationship between *Triakis-Scylliogaleus* and *Furgaleus-Hemitriakis*. Despite variation in their reproductive mechanism, species of *Galeorhinus*

(aplacental viviparity) and *Hypogaleus* (placental viviparous) were showed sister relationship and supported monophyly of Galeorhinini. This study provided evidence of paraphyly of the genera Triakis and Mustelus.

5.5 Mobulidae

In Mobulidae family, species of *Manta* is nested within the genus *Mobula* (Aschliman et al. 2012a, b; Aschliman 2014). Phylogeny analyses by concatenated mitochondrial and nuclear data revealed paraphyly of Mobulidae family and phylogeny tree with three clades. Clade one contained *Mobula japonica*, *M. mobular*, *M. tarapacana*, *Manta birostris* and *M. alfredi*. *Mobula eregoodootenkee* and *M. kuhlii* formed as a second clade. Third clade was composed of *M. munkiana*, *M. hypostoma* and *M. rochebrunei.* Based on genetic distance values and phylogenetic position, *M. mobular* and M. *japonica* could be a single species (Poortvliet et al. 2015).

6 Studies on Trait Evolution and Phylogenetic Diversity

Molecular phylogenetic approach was used to investigate the evolution of bioluminiscence trait in squaliform sharks. This study showed that the photophores might have originated during the transition of the lower to upper Cretaceous time, while families Dalatiidae, Etmopteridae, Oxynotidae and Somniosidae were splitting from their common ancestor (Straube et al. 2015). Chen and Kishino (2015) investigated patterns of phylogenetic diversity (PD) of sharks across the world and advocated to consider PD for formulating conservation measures. Gruber et al. (2016) reported evolution of biofluorescence in families Urotrygonidae, Orectolobidae and Scyliorhinidae using phylogenetic approach.

References

Aschliman NC (2014) Interrelationships of the durophagous stingrays (Batoidea: Myliobatidae). Environ Biol Fish 97:967–979

Aschliman NC, Nishida M, Miya M, Inoue JG, Rosana KM, Naylor GJ (2012a) Body plan convergence in the evolution of skates and rays (Chondrichthyes: Batoidea). Mol Phylogenet Evol 63:28–42

Aschliman NC, Claeson KM, McEachran JD (2012b) Phylogeny of batoidea. In: Carrier JC, Musick JA, Heithaus MR (eds) Biology of sharks and their relatives, 2nd edn. CRC Press, Boca Raton, FL

Bernardi G, Sordino P, Powers DA (1992) Nucleotide sequence of the 18S ribosomal ribonucleic acid gene from two teleosts and two sharks and their molecular phylogeny. Mol Mar Biol Biotechnol 3:187–194

Chen H, Kishino H (2015) Global pattern of phylogenetic species composition of shark and its conservation priority. Ecol Evol 5:4455–4465
Compagno LJV (1973a) Interrelationships of living elasmobranchs. Zool J Linnean Soc 53:15–61
Compagno LJV (1973b) Phyletic relationships of living sharks and rays. Am Zool 17:303–322
Compagno LJV (1977) Phyletic relationships of living sharks and rays. Am Zool 17:303–322
Compagno LJV (1988) Sharks of the order Carcharhiniformes. Princeton University Press, Princeton, NJ, pp 1–542
Compagno LJV (1999) Checklist of living elasmobranchs. In: Hamlett WC (ed) Sharks, skates, and rays: the biology of elasmobranch fishes. Johns Hopkins University Press, Baltimore, pp 471–498
Compagno LJV (2001) Sharks of the world. An annotated and illustrated catalogue of shark species known to date, vol 2. Bullhead, mackerel and carpet sharks (Heterodontiformes, Lamniformes and Orectolobiformes). FAO, Rome
Compagno LJV (2005) Checklist of living chondrichthyan fishes. In: Fowler SL, Cavanagh RD, Camhi M, Burgess GH, Caillet GM, Fordham SV, Simpfendorfer CA, Musick JA (eds) Sharks, rays, and chimaeras: the status of the Chondrichthyan fishes. Status survey. IUCN, Gland, pp 401–423
Corrigan S, Beheregaray LB (2009) A recent shark radiation: molecular phylogeny, biogeography and speciation of wobbegong sharks (family: Orectolobidae). Mol Phylogenet Evol 52:205–216
de Carvalho MR (1996) Higher-level elasmobranch phylogeny, basal squaleans, and paraphyly. In: Stiassny MLJ, Parenti LR, Johnson GD (eds) Interrelationships of fishes. Academic, San Diego, pp 35–62
de Carvalho MR, Maisey JG (1996) Phylogenetic relationships of the Late Jurassic shark Protospinax Woodward, 1919 (Chondrichthyes: Elasmobranchii). In: Arratia G, Viohl G (eds) Mesozoic fishes: systematics and paleoecology. Verlag Dr. Friedrich Pfiel, Munich, pp 9–46
Dingerkus G (1986) Interrelationships of orectolobiform sharks (Chondrichthyes: Selachii). In: Proceedings of the 2nd international conference on Indo-Pacific fishes, Tokyo, pp 227–245
Dosay-Akbulut M (2008) The phylogenetic relationship within the genus Carcharhinus. C R Biol 331:500–509
Douady CJ, Dosay M, Shivji MS, Stanhope MJ (2003) Molecular phylogenetic evidence refuting the hypothesis of Batoidea (rays and skates) as derived sharks. Mol Phylogenet Evol 26:215–221
Dunn KA, McEachran JD, Honeycutt RL (2003) Molecular phylogenetics of myliobatiform fishes (Chondrichthyes: Myliobatiformes), with comments on the effects of missing data on parsimony and likelihood. Mol Phylogenet Evol 27:259–270
Gilbert CR (1967) A revision of the hammerhead sharks (Family Sphyrnidae). Proc US Natl Mus 119:1–87
Goto T (2001) Comparative anatomy, phylogeny, and cladistic classification of the order Orectolobiformes (Chondrichthyes: Elasmobranchii). Mem Grad Sch Fish Sci Hokkaido Univ 48:1–100
Grogan E, Lund R (2004) The origin and relationships of early chondrichthyes. In: Carrier J, Musick J, Heithaus M (eds) Biology of sharks and their relatives. CRC, Washington, DC, pp 3–31
Gruber DF et al (2016) Biofluorescence in Catsharks (Scyliorhinidae): fundamental description and relevance for elasmobranch visual ecology. Sci Rep 6:24751. https://doi.org/10.1038/srep24751
Heemstra PC, Smith MM (1980) Hexatrygonidae, a new family of stingrays (Myliobatiformes: Batoidea) from South Africa, with comments on the classification of batoid fishes. Ichthyol Bull J.L.B. Smith Inst Ichthyol 43:1–17
Heinicke MP, Naylor GJP, Hedges SB (2009) Cartilaginous fishes (Chondrichthyes). In: Hedges SB, Kumar S (eds) The timetree of life. Oxford University Press, New York, p 320
Human BA, Owen EP, Compagno LJV, Harley EH (2006) Testing morphologically based phylogenetic theories within the cartilaginous fishes with molecular data, with special reference to the catshark family (Chondrichthyes; Scyliorhinidae) and the interrelationships within them. Mol Phylogenet Evol 39:384–391

Iglésias SP, Lecointre G, Sellos DY (2005) Extensive paraphylies within sharks of the order Carcharhiniformes inferred from nuclear and mitochondrial genes. Mol Phylogenet Evol 34:569–583

IUCN (2016) The IUCN red list of threatened Species version 2016-3. http://www.iucnredlist.org. Downloaded on 15 Mar 2017

Kriwet J (2004) The systematic position of the Cretaceous sclerorhynchid sawfishes (Elasmobranchii, Pristiorajea). In: Arriata G, Tintori A (eds) Mesozoic fishes 3. München, Verlag Dr Friedrich Pfiel, pp 57–73

Lawson R, Burch SJ, Oughterson SM, Heath S, Davies DH (1995) Evolutionary relationships of cartilaginous fishes: an immunological study. J Zool 237:101–106

Lim DD, Motta P, Mara K, Martin AP (2010) Phylogeny of hammerhead sharks (Family Sphyrnidae) inferred from mitochondrial and nuclear genes. Mol Phylogenet Evol 55:572–579

Lopez JA, Ryburn JA, Fedrigo O, Naylor GJ (2006) Phylogeny of sharks of the family Triakidae (Carcharhiniformes) and its implications for the evolution of carcharhiniform placental viviparity. Mol Phylogenet Evol 40:50–60

Lovejoy NR (1996) Systematics of myliobatoid elasmobranchs: with emphasis on the phylogeny and historical biogeography of neotropical freshwater stingrays (Potamotrygonidae: Rajiformes). Zool J Linnean Soc 117:207–257

Maisey JG (1980) An evaluation of jaw suspension in sharks. Am Mus Novit 2706:1–17

Mara KR, Motto PJ, Martin AP, Hueter RE (2015) Constructional morphology within the head of hammerhead sharks (sphyrnidae). J Morphol 276:526–539

Marra JN, Richards VP, Early A, Bogdanowicz SM, Pavinski Bitar PD, Stanhope MJ, Shivji SM (2017) Comparative transcriptomics of elasmobranchs and teleosts highlight important processes in adaptive immunity and regional endothermy. BMC Genomics 18:87

McEachran JD, Aschliman N (2004) Phylogeny of batoidea. In: Carrier JC, Musick JA, Heithaus MR (eds) Biology of sharks and their relatives. CRC, Boca Raton, FL, pp 79–113

McEachran JD, Dunn KA (1998) Phylogenetic analysis of skates, a morphologically conservative clade of elasmobranchs (Chondrichthyes: Rajidae). Copeia 2:271–290

McEachran JD, Dunn KA, Miyake T (1996) Interrelationships of the batoid fishes (Chondrichthyes: Batoidea). In: Stiassny MJ, Parenti LR, Johnson GD (eds) Interrelationships of fishes. Academic Press, London, pp 63–82

Naylor GJP, Ryburn JA, Fedrigo O, López JA (2005) Phylogenetic relationships among the major lineages of modern elasmobranchs. In: Hamlett WC, Jamieson BGM (eds) Reproductive biology and phylogeny, vol 3. Science, Enfield, NH, pp 1–25

Naylor GJP, Caira JN, Jensen K, Rosana KAM, Straube N, Lakner C (2012) Elasmobranch phylogeny: a mitochondrial estimate based on 595 species. In: Carrier JC, Musick JA, Heithaus MR (eds) Biology of sharks and their relatives. CRC, Boca Raton, pp 31–56

Nelson JS (2006) Fishes of the world, 4th edn. Wiley, Hoboken, NJ. 601 p

Nishida K (1990) Phylogeny of the suborder Myliobatoidei. Mem Fac Fish Hokkaido Univ 37:1–108

Pavan-Kumar A, Gireesh-Babu P, Suresh Babu PP, Jaiswar AK et al (2014) Molecular phylogeny of elasmobranchs inferred from mitochondrial and nuclear markers. Mol Biol Rep 41:447–457

Peach MB, Rouse GW (2004) Phylogenetic trends in the abundance and distribution of pit organs of elasmobranchs. Acta Zool 85:233–244

Poortvliet M, Olsen JL, Croll DA, Bernardi G, Newton K, Kollias S, O'Sullivan J, Fernando D, Stevens G et al (2015) A dated molecular phylogeny of manta and devil rays (Mobulidae) based on mitogenome and nuclear sequences. Mol Phylogenet Evol 83:72–85

Regan CT (1906) A classification of the selachian fishes. Proc Zool Soc Lond 1906:722–758

Rocco L, Liguori I, Costagliola D, Morescalchi MA, Tinti F, Stingo V (2007) Molecular and karyological aspects of Batoidea (Chondrichthyes, Elasmobranchi) phylogeny. Gene 389:80–86

Shirai S (1992) Squalean phylogeny: a new framework of squaloid sharks and related taxa. Hokkaido University Press, Sapporo

Shirai S (1996) Phylogenetic interrelationships of neoselachians (Chondrichthyes: uselachii). In: Stiassny MLJ, Parenti LR, Johnson GD (eds) Interrelationships of fishes. Academic, San Diego, pp 9–34

Straube N, Li C, Claes JM, Corrigan S, Naylor GJP (2015) Molecular phylogeny of Squaliformes and first occurrence of bioluminescence in sharks. BMC Evol Biol 15:62

Vélez-Zuazo X, Agnarsson I (2011) Shark tales: a molecular species-level phylogeny of sharks (Selachimorpha, Chondrichthyes). Mol Phylogenet Evol 58:207–217

Venkatesh B, Kirkness EF, Loh YH, Halpern AL, Lee AP et al (2007) Survey sequencing and comparative analysis of the elephant shark (*Callorhinchus milii*) genome. PLoS Biol 5(4):e101. https://doi.org/10.1371/journal.pbio.0050101

White EG (1937) Interrelationships of the elasmobranchs with a key to the order Galea. Bull Am Mus Nat Hist 74:25–138

Winchell CJ, Martin AP, Mallatt J (2004) Phylogeny of elasmobranchs based on LSU and SSU ribosomal RNA genes. Mol Phylogenet Evol 31:214–224

A Review on DNA Barcoding on Fish Taxonomy in India

V. Sachithanandam and P. M. Mohan

Abstract DNA barcoding has been advanced as an efficient tool to develop for species identification and discovery through the use of short, standardized mitochondrial cytochrome c oxidase I (COI) gene region. Fishes are a highly diverse group of vertebrates; the identification of fish species through DNA Barcoding will provide new perspectives in ecology and systematics of fish taxonomy. Despite extensive traditional taxonomic studies, for a variety of reasons the identification of fishes can be problematic and time taking process, even for experts. In these contexts, DNA barcoding tool is proving to be useful information on species identification in gene level. In connection with this, the Fish Barcode of Life campaign (FISH-BOL), an international research collaboration centre, was established for all fishes' DNA barcode library for reference sequences and to monitor the DNA barcode project progress in regional level. The DNA barcode sequence from any specimen's fish tissue, fin, egg or larva can be matched against these reference sequences retrieved from NCBI and BOLD system for species discrimination/identification. This chapter's aim was to investigate current status of fish barcode, approaches and future direction of DNA barcoding in fisheries sciences. Further, the current status of barcoding studies with reference to fish taxonomy in India is evaluated and a detailed review on existing literature is carried out at worldwide, national and regional level. The study results elucidated that marine invertebrates' DNA barcode study is at infancy stage in India.

Keywords Biodiversity · COI · Cytochrome c oxidase · DNA barcoding · Species identification

V. Sachithanandam (✉)
National Centre for Sustainable Coastal Management, Ministry of Environment, Forests & Climate Change, Anna University, Chennai, India

Department of Ocean Studies and Marine Biology, Pondicherry University, Port Blair, Andaman and Nicobar Island, India

P. Mohan
Department of Ocean Studies and Marine Biology, Pondicherry University, Port Blair, Andaman and Nicobar Island, India

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_16

1 Background

The major part of earth's biodiversity is yet to be explored. There are approximately 1.7 million named species, and biologists have estimated that some 8 million species have not yet been described (Stoeckle and Hebert 2008). It is evident there is an erosion of earth's biodiversity and that one-third of the inhabitants are facing extinction in the coming decades. DNA barcoding is a technique used for identifying organisms based on short standardized fragments of genomic DNA. This will speed up the cataloguing for millions of species yet to be named. For animals, mitochondrial genes like 16S, COI, etc. are ideal gene targets to be used as DNA barcodes (Saccone et al. 1999). Currently, 430,000 barcodes representing about 50,000 species (30% of all known species) have been collected (Silva-Brandao et al. 2009; International Barcode of Life 2010). Fishes are a highly diverse group of vertebrates; the identification of fish species through DNA barcoding will provide new perspectives in ecology and systematics of fishes (Hubert et al. 2008; Ward et al. 2005, 2009). There are approximately 32,100 extant fish species worldwide (as on Aug, 2011, www.Fishbase.org), which constitutes about 50% of all vertebrates. DNA barcodes have been obtained for 6000 species of fishes, including 400 species from New Zealand, 207 species from Australia, 250 species from South Africa and 100 species from Pacific Canada (Ward et al. 2009). Barcoding (cytochrome C oxidase type I (COI) gene) is a major field of research in fish taxonomy, and it is useful for the characterization of fishes existing throughout world. Fish DNA barcoding, based on the sequencing of a uniform area of COI, has received significant interest as an accurate tool for species identification, authentication and phylogenetic analysis, which is followed worldwide. DNA barcoding is a method, based on short standardized gene sequences of DNA (Hebert et al. 2003a), used for species identification. It has been used inter alia for identification of fish species (Hebert et al. 2004b, 2008; Ward et al. 2005; Persis et al. 2009; Steinke et al. 2009a; Lakra et al. 2010; Sachithanandam et al. 2011, 2012). The effectiveness of barcoding has been demonstrated in diverse taxa, including butterflies (Hebert et al. 2003a, b), birds (Hebert et al. 2004a, b), fishes (Ward et al. 2005; Persis et al. 2009; Steinke et al. 2009a; Lakra et al. 2010; Sachithanandam et al. 2011, 2012), sponges, bivalves (Jarnegren et al. 2007) and mammals (Clare et al. 2007). DNA barcoding is being applied in the field of fish conservation (Holmes et al. 2009), fisheries management aspects (Rasmussen et al. 2009) and food safety analysis, with mislabelling being revealed through COI sequences (Wong and Hanner 2008).

In 2005 FISH-BOL (Fish Barcode of Life Initiative) was set up to establish a library of COI gene sequences of all fish species, to enable global taxonomic identification (Ward et al. 2009; Eschmeyer 2010). FISH-BOL's primary work is led by a research team with responsibility for collection of samples, traditional identification using morphological keys and COI gene sequencing of species in their geographic regions. Current progress of DNA barcoding in the Arctic and Antarctic regions is 74% and 50%, respectively. Other regions from tropical to subtropical show good progress, with 20% coverage from Australia, Mesoamerica

continental and Oceania. However, species-rich regions of Asia, South America and Africa show low progress (Becker et al. 2011). In India, 11,023 fish species have been morphologically reported (Nelson 2006; Mecklenburg et al. 2011). Reports suggest that only 1918 of these (17.4%) have been barcoded (Becker et al. 2011). The present article aim was to investigate global status, approaches and future direction of DNA barcoding in fisheries sciences. A detailed review was carried out on existing literature was dispensed in worldwide, national and regional level in India.

2 DNA Barcoding in Fishes and Other Animals: Status

Taxonomy is the science assigning names to species and higher taxa. The world is facing a global biodiversity crisis (Novacek and Cleland 2001; Bellwood et al. 2004). The rapid loss of marine and terrestrial biodiversity has prompted efforts to catalogue this biodiversity in Census of Marine Life (CoML). Nowadays morphologically taxonomic identifications are declining in number the reason behind this being the decline of taxonomic skills (Hopkins and Freckleton 2002) and insufficient funding. Besides, the upcoming scientists focus on the advance techniques of molecular methods. Hebert et al. (2010) proposed the molecular technique DNA barcoding using a short DNA sequence has been increasingly adopted in taxonomic identification studies. Currently, these methods are being applied to a wide range of taxonomic groups (Hebert et al. 2004a, b; Ward et al. 2005; Meyer et al. 2008). The main goal of these DNA barcoding is to characterize the biological diversity using a short DNA sequence that will facilitate and expedite taxonomic identification. This initiative has helped in the discovery of new species many of which have been shown to be morphologically cryptic, thereby considerably improving biodiversity assessment in poorly studied benthic fauna (Blaxter et al. 2004; Derycke et al. 2005; Leasi and Todaro 2009). The studied literature revealed that many works have been undertaken using DNA barcoding tools and sole source to identify many species and sibling species. Globally there are different molecular markers that will help unravel the evolutionary history of annelids, mainly 18S nuclear ribosome (Colgan et al. 2001; Bleidorn et al. 2006). Envall et al. (2006) used two mitochondrial genes 12S rDNA and 16S rDNA and one nuclear gene 18S rDNA in clitellate. Sjolin et al. (2005) studied the 18S rDNA, and new 16S rDNA sequence data in Tubificidae (Thomas et al. 2005; Satheeshkumar and Jagadeesan 2010; Erseus et al. 2010) used 16S mitochondrial gene sequence.

DNA barcoding is a well-accepted taxonomic method which uses a short genetic marker to facilitate identification of a particular species even by non-specialist. COI has been accepted as universal barcode to delineate animal life. At present the numbers of taxonomists are declined, and there are insufficient numbers of specialists (Rubinoff 2006; Hebert et al. 2003a; Packer et al. 2009) in this field to handle the present taxonomy scenario. Hebert et al. (2003b) estimated that there are 15,000 taxonomists which will be required for the identification of 10–15 million species. It

will cumbersome to identify the new species, which is having maximum similarities in morphological and environment identities. In these circumstances, the modern tool of DNA Barcoding evolved and provides a better understanding of morphological classifications, which were appeared due to the slight variation in the environment or other factors (Hebert et al. 2003a, b, 2004a; Ward et al. 2005).

Barcoding is a compliment to current research in taxonomic studies by providing detailed information which will be helpful for the identification of taxa in nature (Hajibabaei et al. 2006a, 2007a; Hebert et al. 2004b; Ball et al. 2005; Saunders 2005; Ward et al. 2005). Because the traditional taxonomy was able to provide descriptions for less than quarter of the world species using variety of morphological keys as a tool (Packer et al. 2009), a faster and reliable tool is essential. In fact, DNA-based identification cannot be accomplished without the involvement and expertise of taxonomists who identify specimens from which reference sequences were obtained (Ball et al. 2005; Hebert et al. 2003a; Hebert and Gregory 2005; Coyne and Orr 2004). Hubert et al. (2008) explained the efficiency of the DNA barcoding hinges on the degree of sequence divergence among species and species-level identifications. Further, Escalante et al. (2011) and Baird and Sweeney (2011), COI gene sequence is used for the bio-monitoring in benthology. Sweeney et al. (2011) explain that certain cases, where ambiguity existed their morphological data need to be re-examined using the additional molecular data for species clarification. Min and Hickey (2007a) proved that barcoding was sufficient for species identification among the fungi. DNA barcoding has a new milestone achievement of mini-barcoding studies which was carried out by Meusnier et al. (2008). The systems of classical taxonomic identification are carried out solely by morphology whereas it fails to identify its characters for some groups of organisms, the DNA barcoding is considered an ideal tool for species identification (Mitchell 2008; Pages et al. 2009; Zettler et al. 2002; Abriouel et al. 2008; Hebert et al. 2003a). A detailed gene and gene classification used for species taxonomy were reviewed and provided in Table 1.

Mitochondrial DNA (mtDNA) profiling studies are carried out in the cytoplasmic mtDNA which is inherited from the female parent so that each copy is identical. It offered the insights for the population structure and greatly contributed to the establishment of phylogeographic information (Avise et al. 1987). The development of "Universal" primers (Kocher et al. 1989) made possible DNA amplification of COI gene by using PCR and their direct sequencing for many number of phyla.

Also, the sequences of mtDNA diverged quickly by its information, and the gene order with the composition is relatively uniform (Simon et al. 1994, 2006). When compared to the nuclear genome, the mitochondrial genome lacks introns, which restricted to have an exposure for recombination and has a haploid mode of inheritance (Saccone et al. 1999) is also an added advantage for the DNA barcoding tool.

Mitochondrial genes are useful to study the species that diverged recently because they have a high rate of substitution. However, if the divergence event was not recent, nuclear genes are ideal for phylogenetic analysis (Lin and Danforth 2004). mtDNA sequence divergences had been successfully used to discriminate between species of

Table 1 Synoptic view of DNA barcoding which employed COI gene to identify animals taxa and its K2P values

Sl. No.	References	Taxa name and no.	% of the identified	K2P % intraspecific vs. interspecific	Remark
1	Hebert et al. (2003a)	2238 Annelida, Anthropoda, Chordata, 4, Echinodermata, Mollusca, Nematoda, Platyhelminthes	>98	Overall <2 vs. 11.3	The efficacy of COI in identifying species from eight major and several minor phyla plus a variety of arthropod classes assessed. Cnidarians showed less COI variation between species then all other taxonomic groups, 94.1% vs. 1.9% showing <2 K2P between spp. ($p < 0.0001$)
2	Remigio and Hebert (2003)	70 Gastropod spp.	98		DNA barcoded to identify species and insertion or deletions more common in COI in this taxonomic group
3	Hebert et al. (2004b)	260 Avian spp.	100	0.43 vs. 7.93	Four possible cryptic species identified
4	Penton et al. (2004)	2 Daphnia spp.	100		Identification of morphologically cryptic species with overlapping distribution (Crustacean)
5	Hebert and Barrett (2005)	203 Arachnid spp.	100	1.4 vs. 16.4	Mean intra-and interspecific nucleotide divergences did not overlap expect
6	Ward et al. (2005)	207 Marine fishes from Australia	100	0.39 vs. 9.93	Efficacy of COI at identifying species and taxonomic relationships assessed
7	Dooh et al. (2006)	4 Crustacean spp.	96–98	1.5 vs. 27	Using barcodes to examine the phycology of two glacial relict crustacean taxa in North America
8	Hajibabaei et al. (2006a)	Lepidopteran and	100	0.1–0.4 vs. 5.4–8.9	Using mini-barcodes to identify species

(continued)

Table 1 (continued)

Sl. No.	References	Taxa name and no.	% of the identified	K2P % intraspecific vs. interspecific	Remark
9	Hajibabaei et al. (2006b)	521 Lepidopteron spp.	97.90	0.17–0.46 vs. 4–6	Morphologically distinct sympatric species from three families identified
10	Costa et al. (2007)				DNA barcoding of Crustacean species
11	Elias-Gutierrez et al. (2008)	2 Cladoceran spp.	96–98	14.3	A new cryptic species (Crustacea, Chydoridae) from the desert region of Mexican
12	Moura et al. (2008)	2 Genera of elasmobranchs	97		DNA barcoding used to resolve within genera identification problems in deep water sharks
13	Holmes et al. (2009)	20 Shark spp. 7 ray spp.	91.5		Identifying shark species from dried fins for conservation purposes
14	Radulovici et al. (2010)	Marine crustaceans from the Gulf of St Lawrence	95–97		COI gene used to identify 80 Mollusca species
15	Steinke et al. (2009a)	391 Fish spp.	100	0.42 vs. 10.81	Producing a DNA database of ornamental fish from Pacific Canada
16	Vargas et al. (2009)	5 Sea turtle spp.	100		DNA barcoding of Brazilian sea turtles

North American birds (Hebert et al. 2004b), spiders (Hebert and Barrett 2005), cryptic species of butterflies (Hebert et al. 2004a), mosquitoes (Besansky et al. 2003), leeches (Siddall and Budinoff 2005), springtails (Hogg and Hebert 2004), beetles (Monaghan et al. 2005), oligochaetes (Nylander et al. 1999), naidid worms (Bely and Wray 2004), extinct moas (Lambert et al. 2005) and various other species of vertebrates and invertebrates (Saccone et al. 1999; Hebert et al. 2003b).

The barcode system was based on sequence diversity in a single gene (Schander and Willassen 2005) region, i.e., a section of the mitochondrial DNA cytochrome C oxidase I gene (COI). These sequences demonstrated higher order relationship than the shallower divergence (Seifert et al. 2007; Clare et al. 2008). This technique provides higher flexibility for the identification of species in large taxonomic assemblages (Caterino and Tishechkin 2006; Pegg et al. 2006; Hajibabaei et al. 2006c, 2007b). The ability to use COI gene to identify species would enable the identification

of cryptic and polymorphic taxa and also identify and associate individuals of life stages other than adult to their correct species (Schander and Willassen 2005; Kartavtsev et al. 2009; Kochzius et al. 2010; Duo et al. 2010). Further, COI locus sequence analysis was an economically feasible option for taxonomic study (Ball et al. 2005). The use of COI barcodes is a powerful tool for species identification of individually isolated fish eggs, larvae and fillets and fins (Ward et al. 2005; Hubert et al. 2008; Steinke et al. 2009b). Nguyen and Seifert (2008) discovered three new species of *Leohumicola* using morphological characters and phylogenetic analyses of DNA barcodes using COI sequences. According to Min and Hickey (2007b), the reducing sequence length had a profound effect on the accuracy of resulting phylogenetic trees, but surprisingly short sequences still yield accurate species identifications.

Radulovici et al. (2010) suggest that DNA barcoding is the emerging tool for marine biodiversity study for identification at the species level. Several studies were attempted in marine polychaetes, and its COI data clearly species discriminated without ambiguity (Meyer et al. 2008). Steinke et al. (2009b) described a software tool for DNA barcoding using distance methods for species identification. Siddall and Budinoff (2005) made an attempt on molecular phylogenetic analysis for the DNA barcoding evidence for widespread introductions of a leech from South America. Further, Halanych and Janosik (2006) work has summarizes the molecular data relevant to the evolutionary history and molecular phylogeny of siboglinid annelids and investigated the evaluations and molecular systematics of polychaetes.

It would be beneficial to have a standard segment (such as the barcoding region) to use for routine identifications. DNA barcodes have potential to be useful for species identification without the aid of a taxonomist in certain situations. Barcoding had been proposed as a quality control measure to confirm the identity of specimens (Mitchell 2008). The commercial uses of barcoding, excited in pest identification, invasive species detection and fishery management were also worth pursuing (Mitchell 2008; Rock et al. 2008).

The fish taxonomy mainly consists of environmental factors and its dynamics of larval dispersal. Many unanswered issues in the ecology and evolution of marine population centre on how far planktonic larvae disperse away from their parents (Levin et al. 2006). Regardless of the importance of the ecological processes affected by larval fish dynamics, the inability of unambiguous taxonomic identification of early life stages of many taxa is still a major burden that impairs the proficient management of these populations. Further, the larval stages of fish diversity studies attended several scientific groups through classical taxonomy approches are facing difficulty in distinguishing genus and species level (Chow and Walsh 1992; Victor et al. 2009).

Despite the great promise of DNA barcoding, it had been controversial in some scientific circles (Will and Rubinoff 2004). Yet, recent results illustrated some straightforward benefits from the use of a standardized molecular approach for identification (Hebert et al. 2003a; Hebert and Gregory 2005). Similarly, intraspecific phenotypic variation often overlapped that of sister taxa in nature, which lead to incorrect identifications, if based on phenotype only (Pfenninger et al. 2006). The

recently introduced next-generation sequencing (NGS) approaches in biodiversity science have the potential to further extend the application of DNA information (Hajibabaei et al. 2011). Further, Ratnasingham and Hebert (2007) reported that DNA technology promised the development of portable devices which will both gather barcode sequences in minutes and use an on-board barcode reference library to generate identifications.

The morphological ambiguity existed for several fish genera/species, and a proper identification has imperative for management and trade (Eschmeyer et al. 1998; Wiens and Servedio 2000; Hebert et al. 2003b; Hebert and Barrett 2005; Hebert and Gregory 2005; Nelson 2006; Ward et al. 2009) which the DNA barcode resolves this problem and also helps to discover new/cryptic species. Further, the morphological-based studies need lot of taxonomist who are dwindling due to nature of hard work (Steinke et al. 2009a). So, the barcode methods may provide a new taxonomist. The barcode studies also provide the understanding of the eco-diversity, especially for the marine fishes which cannot be monitored continuously (Dasmahapatra and Mallet 2006). Moreover, the fisheries managers and scientists are struggling with a lack of basic information for many shark and ray species which can be carried out through barcode of confiscated materials like shark meat, fin, bones, etc., which provide much missed details which can be used for scientific and also for conservation and management (Holmes et al. 2009).

Larval fishes were frequently not identified to species level due to their small size and limited morphological development (Webb et al. 2006; Richardson et al. 2006). These leads to difficulties for to understanding the life histories of fishes, that also marine fishes makes cumbersome the task. Very few species and/or genus only explored for the taxonomic identification, early life history and phylogenetic relationships (perches fishes—*Lutjanus*) which were far from complete and continually reviewed (Rivas 1949; Chow and Walsh 1992; Leis 1986, 2005; Miller and Cribb 2007). Shirak et al. (2009) studied the barcoding and taxonomic analysis of the five *Tilapiine* species. The larvae and newly settled juveniles of the Cubera Snapper were identified by DNA barcoding (Victor et al. 2009). The stomatopod larvae were studied under barcode technique (Barber and Boyce 2006) to understand the biodiversity of gonodactylid (mantis shrimp). Zemlak et al. (2009) had been reported that Indo-Pacific fishes, for a substantial number, were overlooked under the category of commercially important fishes. Moreover, the COI gene sequences for marine fishes were successfully carried out in Australia (Ward et al. 2005), Pacific Canada (Steinke et al. 2009a), North Atlantic (Ward et al. 2008), South Africa (Zemlak et al. 2009), Finland (Salokannel et al. 2010) and Great Barrier Reef (Pegg et al. 2006). The faunal studies progressed by the COI sequence for their species identification/taxonomy were listed out in Table 2.

At present fishes were the most studied marine groups and were currently barcoded within two global campaigns, FISH-BOL (http://www.fishbol.org) and SHARK-BOL (http://www.sharkbol.org) (Ward et al. 2009). Existing fish barcoding results were reviewed, and lacunas were presented in detailed by Ward et al. (2009) Radulovici et al. (2010). According to Johnson and Keener (1984) suggested that gene level species discrimination study needed for the groupers fishes, to increase

Table 2 Identified K2P values for the COI gene and other genes, for species identifications

Sl. No.	Authors and year	Intraspecific K2P values (%)	Inter generic K2P values (%)	Within family and order K2P values (%)	Animal included in DNA barcoding paper tile
1	Hebert et al. (2004b)	mean K2P 0.27 [within species 0.43]	Within genus 7.93	Within family 12.71	Identification of birds through DNA barcode
2	Clare et al. (2007)	Intraspecific variation mean = 0.60.	Congeners mean =7.80	Within family 21.26	DNA barcoding of neotropical bats
3	Ward et al. (2008)	Intraspecific K2P variation 0.028 ($n = 22$) of *Z. faber*; avg. 0.20 to 0.23			DNA barcoding of shared fish species from the North Atlantic and Australasia
4	Steinke et al. (2009b)	Intraspecific variation (mean = 0.21)	Congeneric K2P 10.81		DNA identifications for the ornamental fish trade
5	Khedkar et al. (2009)	Molecular variance analysis revealed that 96%/6 haplotypes per species			DNA barcoding of fish in the Godavari River, India
6	Odeny et al. (2009)	Within species was 0.11			Para taxonomy for fishery surveys using DNA barcoding
7	Oliveira et al. (2009)	Average = 0.6 within species	Average = 8.7 within genera	Average K2P = 17.6	DNA barcode of freshwater fishes of Brazil
8	Rasmussen et al. (2009)	Intraspecies divergences (mean 0.26) range 0.04 to 1.09	Congeneric K2P values of 32-fold greater, at 8.22 (range 3.42–12.67)		Commercially important salmon through DNA barcoding of in North America
9	Zemlak et al. (2009)	K2P distances between African and Australian (mean = 5.10)			DNA barcoding reveals overlooked marine fishes

(continued)

Table 2 (continued)

Sl. No.	Authors and year	Intraspecific K2P values (%)	Inter generic K2P values (%)	Within family and order K2P values (%)	Animal included in DNA barcoding paper tile
10	Aliabadian et al. (2009)	*COI:* intraspecific K2P distances averaged 0.24 (SD = 0.59)	*COI:* Intrageneric K2P 24-fold higher than the mean intraspecific K2P distances	*COI:* Mean K2P within families = 11.46 and orders = 15.80	Molecular identification of birds: performance of distance-based DNA barcoding in three genes
		Cyto b gene: intraspecific K2P averaged 0.74 (SD = 1.21)	*Cyto b gene:* Intrageneric K2P 11-fold higher than the mean intraspecific K2P distances	*Cyto - b:* Mean divergences within families = 13.97 and Order = 19.50	
		16S gene: intraspecific K2P averaged 0.48% (SD = 1.06)	*16S gene:* Intrageneric K2P distances sevenfold higher than the mean intraspecific K2P	*16S gene:* Mean K2P within families = 6.51	
11	Steinke et al. (2009a)	Intraspecific pacific Canada fishes K2P – 0.25	Genera 3.75		Pacific Canada's fishes discrimination using COI gene sequences
12	Kartavtsev et al. (2009)	Average intraspecies, 0.11 ± 0.04	Intra-genus 1.87 ± 0.68	Intra-family 12.67 ± 0.28; Intra-order 16.52 ± 0.10.	Molecular phylogenetics of prickle backs and other percoid fishes from the Sea of Japan
13	Persis et al. (2009)	Mean intraspecific K2P values in family = 0.24	Avg. congeneric K2P = 17.2	mean K2P family = 0.875%	DNA barcoding of carangid fishes from Andhra coast, India
14	Lakra et al. (2010)	Mean K2P 0.30	Within genus mean K2P 6.66	Families mean K2P – 9.91, order mean K2P – 16.00	Indian marine fishes barcoding 115 species,

(continued)

Table 2 (continued)

Sl. No.	Authors and year	Intraspecific K2P values (%)	Inter generic K2P values (%)	Within family and order K2P values (%)	Animal included in DNA barcoding paper tile
					79 genera, 37 families
15	Zhang (2011)	K2P average 0.319% for intra-specific individuals	K2P average 15.742 among congeners		DNA barcoding of marine fishes in China
16	Zhang and Hanner (2011)	among conspecifics K2P average of 0.3%	Genetic distances averaged 17.6% among congeners		DNA barcoding of marine fishes from Japan
17	Hubert et al. (2011, Unpublished data) Identification coral reef fish larvae through DNA barcoding: a test case with the families Acanthuridae and Holocentridae	Within species 0.001–0.004; *S. caudimaculatum* and *S. spiniferum* that diverged only by 0.007 on average, this result reinforces the view that no canonical threshold applies to the frontier separating populations and species in fishes (Hubert et al. 2008)	Among genera from an average of 0.063 in *Myripristis.* 0–0.11 on average for *Acanthurus*		Hubert et al.
18	Pramual et al. (2011)	Intraspecific K2P—0 to 9.28, with a mean of 2.75	Interspecific K2P 0.34 – 16.05		Cryptic biodiversity and phylogenetic relationships revealed by DNA barcoding of oriental black
19	Carolan et al. (2012)	Within species *B. cryptarum, B. lucorum* and *B. magnus* for 0.004, 0.001 and 0.001	Interspecific K2P: 0.033 to 0.044		DNA barcoded of bumblebee species complex, and colour patterns do not diagnose species
20.	Lorz et al. (2012)	within the species of *R. chathamensis* and *R. abyssalis* (K2P = 0.000);	Inter-clade 0.143 –0.370; an overall		Description of a new species Crustacea, Amphipoda

(continued)

Table 2 (continued)

Sl. No.	Authors and year	Intraspecific K2P values (%)	Inter generic K2P values (%)	Within family and order K2P values (%)	Animal included in DNA barcoding paper tile
		24 specimen *R. aculeate* (0.0089); 9 specimen *R. helleri* (K2P = 0.0003); 13 - individual *R. inflata* (K2P = 0.037)	average divergence 0.284		*Rhachotropis* using COI gene sequences
21	Turanov et al. (2012)	0.06% among the species K2P value	Mean K2P within genus 0.37	Within family K2P—11.83; within order—15.22	Molecular phylogenetic study of eelpout fishes from Eastern Seas

the understanding of relationship of phylogenetic within families. In connection with, Maggio et al. (2005) mitochondrial cytochrome b and 16S rDNA sequences analysis carried out using Eastern Atlantic grouper species phylogenetic and cluster relationship observed for species evolutionary tendencies. Craig and Hastings (2007) carried out barcoding to understand the phylogenetic relationships among the perciform tribe Epinephelinae (Serranidae) which had been poorly understood. Smith and Craig (2007) studied the limits and relationships of Serranid and percid fishes using the nucleotide characters. Koedprang et al. (2007) studied the genetic diversity among the grouper species (eight species) using microsatellite marker.

3 Fish Diversity and DNA Barcode Status in India

India has a wider continental shelf with 8000 km length of coastline with high potential of fishery resources. Coastal waters of India are known for their rich and diverse fish species (Venkataraman and Wafar 2005). Worldwide, about 32,000 species of fish have been catalogued, in 6 classes, 62 orders and 540 families (Eschmeyer 2010), accounting for more than half of all vertebrate species. Major contributions to Indian ichthyology have been made by the pioneering work of day (1878); further, Talwar (1990) reported about 2546 fish species belonging to 969 genera, 254 families and 40 orders. Fifty-seven percent of marine fish genera are sharing and common to the Indo-Pacific and Atlantic and Mediterranean regions (Venkataraman and Wafar 2005). Fish diversity in the seas of the Andaman and

Nicobar Islands also is of special interest in terms of marine zoogeography because they lie at the confluence of the Andaman Sea with the Western Pacific and the Indian Ocean (Rajan 2010). A total of about 1485 species of fish under 603 genera belonging to 177 families is represented from these islands, of which 400 species have commercial significance as food. Among the fish species, 1089 (about 73.38%) are recorded from coral reef environment, 277 from mangroves, 152 from seagrass meadow, 23 from freshwater streams and 101 from offshore environment, while 158 species are commonly observed in mangrove, seagrass, coral reefs and offshore ecosystems (Rao 2003; Rajan and Sreeraj 2012). The catalogued endemic biodiversity of the islands comprises 16 species of fish, 31 species of reptile and 8 species of amphibian (Ramakrishna et al. 2010). Tropical fishery in India is so diverse that there are several misnomers among the fishes identified, and certain similar species are recorded as similar/same for even scientific purposes.

An exploratory survey along the south west coastal waters of India revealed the presence of certain new species recorded akin to their nearest generic counterparts but appearing distinct for a classical taxonomist. In Eastern part of Bay of Bengal, surrounded by the Andaman group of islands has rich marine habitants with good diversity of fishes. Further, it is an isolated, virgin area in terms of exploration as well as exploitation of fishery resources. Indian Marine fishes were attempted for DNA barcoding in East Coast as well as west coastal region. These barcode results reported that the average (K2P) distances within species, genera, families and orders were 0.30%, 6.60%, 9.91% and 16.00%, respectively (Lakra et al. 2010). Persis et al. (2009) studied the Carangid fishes from Kakinada coastal region based on COI gene sequence-based approach. Jhon et al. (2010) studied *Stolephorus* spp. using mtDNA COI method for species identification from Parangipettai coastal areas. Sachithanandam et al. (2011, 2012) a pilot study on DNA barcode in coral reefs fishes especially grouper of Andaman Islands data suggested that ANI fishes has high genetic diversity when compared with mainland coastal region. Govindaraju and Jayasankar (2004) used RAPD for to discriminate grouper fishes species. Jayasankar et al. (2004) carried out a complete approach to Indian mackerel in the East and West Coasts of India through the RAPD techniques, and study revealed that significant genetic difference was observed. Later, Lakra et al. (2007) studied five Indian Sciaenids using the DNA (RAPD) markers.

DNA barcoding of *Lates calcarifer* was studied in Porto Nova coastal region using COI gene for the phylogenetic analysis and GC content and genetic distances compared with worldwide species (Jhon et al. 2010). Kumar et al. (2011) worked in cryptic species of Mugilidae fishes using DNA barcoding tools for phylogenetic and haplotype diversity as revealed by intraspecies genetic distances. Sachithanandam et al. (2011, 2012) brought out barcode details for the grouper subfamily Epinephelinae species of Andaman waters. The studied literature suggested that the monophyly of fishes was not carried out in systematic in India (Sachithanandam et al. 2012). So, the present review article aim was to investigate global status, approaches and future direction of DNA barcoding in fisheries sciences, and existing literature were carried out in worldwide, national and regional level in India coastal region (Tables 1 and 2).

4 DNA Barcoding Emphases

Applications of DNA barcoding tools are emerging in the fields of fish conservation and management aspects such as quota, by-catch monitoring and sustainable fisheries monitoring science (Holmes et al. 2009; Steinke et al. 2009a, b; Rasmussen et al. 2009). In the fields of food safety aspects, DNA barcoding has demonstrated that 25% of fish samples from markets and restaurants in the USA and Canada that were mislabelled or substituted were observed (Wong and Hanner 2008). DNA barcoding can also be applied successfully to cooked or processed fish meat (Smith et al. 2008), grilled or deep-fried fillets (Wong and Hanner 2008) and boiled samples (Shirak et al. 2009) and canning sample use of shorter fragments called mini-barcodes (Hajibabaei et al. 2006a; Meusnier et al. 2008; Rasmussen et al. 2009; Ward et al. 2009).

5 DNA Barcoding Progress

FISH-BOL has the primary goal of gathering DNA barcode records for all the world's fishes, about 32,500 species (Ward et al. 2009; Eschmeyer 2010). The aim of the FISH-BOL organization is to develop COI gene sequences all over the world. But these efforts are reflected in a high coverage from Arctic (74%) and Antarctic (50%). Other regions such as Australia, America and Oceania showed good progress with coverage near about 20%. However, extremely species-rich regions such as Asia and Africa showed lower progress. This might be explained with an observed bias toward the processing of marine species. The FISH-BOL database recored about 7800 species barcoded, among the 5200 species (about 73 percentage) are belonging to the marine (Eschmeyer 2010). In Indian fish fauna of 11,023 species morphologically reported but 1918 species were barcoded around 17.4% progressed reported by Mecklenburg et al. (2011). Therefore, sampling in the coming years should focus on the collection of marine and freshwater species in India and Southeast Asia. Becker et al. (2011) reported that India water has a large potentical resources for barcoding study which are neglected families by morphologically ambiguities.

6 DNA Barcoding Success Rate

The investigation on marine fish specimens identification using DNA barcode tool was sucessfully species discriminated upto 98% identity from Austalian water (Ward et al. 2009). The last decade, DNA barcoding tools is working very effectively for species identification with out any ambiguity. Further, DNA barcoding COI gene

sequences produced regional genetic differentiation and shared haplotypes genetic differences due to the different habitants (Hubert et al. 2008; Ward et al. 2009; Sachithanandam et al. 2012).

7 Some Limitation and Cautions

Ward et al. (2009) analysed DNA barcoding sequences in fishes and birds and opined that the success for barcoding depends upon recent speciation, incorrect taxonomy or hybridization, where barcoding could not differentiate between species. There are many drawbacks to the use of barcoding for species identification, so the scientific community must be cautious in accepting it. Some biological phenomena that potentially interfere with barcoding are heteroplasmy, hybridization, paternal leakage, introgression, polyploidization, recent speciation, incomplete lineage sorting, error in specimen identification and incorrect taxonomy above phenomena which are known to occur to different degrees depending on the data set (Hebert et al. 2003b, 2004a; Mitchell 2008; Ward et al. 2009; Rubinoff 2006; Rock et al. 2008; Langhoff et al. 2009). So, it is essential to have more number of data on individual species, and correct identification of species through traditional morphology and uploading of correct sequence for right species are highly essential to have a best barcoding technology for species identification.

Acknowledgements This work was supported by a Ministry of Earth Sciences (MoES), New Delhi, India, under benthic programme and the higher authorities of Pondicherry University, Puducherry, India, for which the authors are grateful.

References

Abriouel HPAM, Maqueda M, Valdivia E, Bueno MM (2008) Biodiversity of the microbial community in a Spanish farmhouse cheese as revealed by culture-dependent and culture-independent methods. Int J Food Microbiol 127(3):200–208

Aliabadian M, Kaboli M, Nijman V, Vences M (2009) Molecular identification of birds: performance of distance-based DNA barcoding in three genes to delimit parapatric species. PLoS One 4(1):e4119. https://doi.org/10.1371/journal.pone.0004119

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu Rev Ecol Syst 18:489–522

Baird DJ, Sweeney WB (2011) Applying DNA barcoding in benthology: the state of the science. J N Am Benthol Soc 30(1):122–124

Ball SL, Hebert PDN, Burian SK, Webb JM (2005) Biological identification of mayflies (Ephemeroptera) using DNA barcodes. J N Am Benthol Soc 24:508–524

Barber P, Boyce SL (2006) Estimating diversity of Indo-Pacific coral reef stomatopods through DNA barcoding of stomatopod larvae. Proc R Soc Biol Sci 273:2053–2064. https://doi.org/10.1098/rspb

Becker S, Hanner R, Steinke D (2011) Five years of FISH-BOL: brief status report. Mitochond DNA 22(S1):3–9. https://doi.org/10.3109/19401736.2010.535528

Bellwood DR, Hughes TP, Folke C, Nystrom M (2004) Confronting the coral reef crisis. Nature 429:827–833

Bely AE, Wray GA (2004) Molecular phylogeny of naidid worms (Annelida: Citellata) based of cytochrome oxidase I. Mol Phylogenet Evol 30:50–63

Besansky NJ, Severson DW, Ferdig MT (2003) DNA barcoding of parasites and invertebrate disease vectors: what you don't know can hurt you. Trends Parasitol 19(12):545–546

Blaxter M, Elsworth B, Daub J (2004) DNA taxonomy of a neglected animal phylum: an unexpected diversity of tardigrades. Proc R Soc Lond Biol Sci 271:S189–S192

Bleidorn C, Podsiadlowski L, Bartolomaeus T (2006) The complete mitochondrial genome of the orbiniid polychaete *Orbinia latreillii* (Annelida Orbiniidae) – a novel gene order for Annelida and implications for annelid phylogeny. Gene 370:96–103

Carolan JC, Murray TE, Fitzpatrick U, Crossley J, Schmidt H (2012) Colour patterns do not diagnose species: quantitative evaluation of a DNA barcoded cryptic bumblebee complex. PLoS One 7(1):e29251. https://doi.org/10.1371/journal.pone.0029251

Caterino MS, Tishechkin AK (2006) DNA identification and morphological description of the first confirmed larvae of Hetaeriinae (Coleoptera: Histeridae). Syst Entomol 31:405–418

Chow S, Walsh PJ (1992) Biochemical and morphometric analyses for phylogenetic relationships between seven snapper species (Subfamily Lutjanidae) of the Western Atlantic. Bull Mar Sci 50:508–519

Clare EL, Lim BK, Engstrom MD, Eger JL, Hebert PDN (2007) DNA barcoding of Neo tropical bats: species identification and discovery within Guyana. Mol Ecol Notes **7**:184–190. https://doi.org/10.1111/j.1471-8286.2006.01657.x

Clare EL, Kerr KC, Von- Konigslow TE, Wilson JJ, Hebert PDN (2008) Diagnosing mitochondrial DNA diversity: applications of a sentinel gene approach. J Mol Evol 66:362–367. https://doi.org/10.1007/s00239-008-9088-2

Colgan DJ, Hutchings P, Brown S (2001) Phylogenetic relationships within the Terebellomorpha. J Mar Biol Ass UK 81:765–773

Costa F, DeWaard JR, Boutilier J, Ratnasingham S, Dooh RT, Hajibabaei M, Hebert PDN (2007) Biological identifications through DNA barcodes: the case of the Crustacea. Can J Fish Aquat Sci 64(2):272–295

Coyne JA, Orr HA (2004) Speciation. Sinauer Associates, Sunderland, MA

Craig MT, Hastings PA (2007) A molecular phylogeny of the groupers of the subfamily Epinephelinae (Serranidae) with a revised classification of the Epinephelinae. Ichthyol Res 54:1–17. https://doi.org/10.1007/s10228-006-0367-x

Dasmahapatra KK, Mallet J (2006) DNA barcodes: recent successes and future prospects. Heredity, pp 1–2. doi:https://doi.org/10.1038/sj.hdy.6800858

Day F (1878) The fishes of India; being a natural history of the fishes known to inhabit the seas and fresh waters of India, Burma, and Ceylon, Parts 1–4. William Dawson, London, p 778

Derycke S, Remerie T, Vierstraete A, Backeljau T, Vanfleteren J, Vincx M, Moens T (2005) Mitochondrial DNA variation and cryptic speciation within the free-living marine nematode Pellioditis marina. Mar Ecol Prog Ser 300:91–103

Dooh RT, Adamowicz SJ, Hebert PDN (2006) Comparative phylogeography of two North American 'glacial relict' crustaceans. Mol Ecol 15(14):4459–4475

Duo WZ, Yusong G, Wei T, Lu L, EnPu T, ChuWu L, Yun L (2010) DNA barcoding, phylogenetic relationships and speciation of snappers (genus *Lutjanus*). Sci China Life Sci 53(8):1025–1030. https://doi.org/10.1007/s11427-010-4034-0

Elias-Gutierrez M, Jeronimo FM, Iavanova NV, Valdes-Moreno M, Hebert PDN (2008) DNA barcodes for *Cladocera* and *Copepoda* from Mexico and Guantemala, highlights and new discoveries. Zootaxa 1839:1–42

Envall I, Källersj M, Erseus C (2006) Molecular evidence for the non-monophyletic status of Naidinae (Annelida:Clitellata:TubiWcidae). Mol Phylogenet Evol 40:570–584

Erseus C, Rota E, Matamoros L, De Wit P (2010) Molecular phylogeny of Enchytraeidae (Annelida: Clitellata). Mol Phylogenet Evol 57:849–858

Escalante P, Vazquez AI, Escobar PR (2011) Tropical montanenymphalids in Mexico: DNA barcodes reveal greater Diversity. Mitochond DNA 21(S1):30–37
Eschmeyer WN (2010) Catalog of fishes. http://research.calacademy.org/research/ichthyology/catalog/speciesby family.asp
Eschmeyer WN, Ferraris JCJ, Hoang MD, Long DJ (1998) Part 1. Species of fishes. In: Eschmeyer WN (ed) Catalog of fishes. California Academy of Sciences, San Francisco
Govindaraju GS, Jayasankar P (2004) Taxonomic relationship among seven species of groupers (Genus *Epinephelus*, Family Serranidae) as revealed by RAPD fingerprinting. Mar Biotechnol 6:229–237. https://doi.org/10.1007/s10126-003-0021-9
Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006a) DNA barcodes distinguish species of tropical Lepidoptera. Proc Natl Acad Sci U S A 103:968–971
Hajibabaei M, Smith MA, Janzen DH, Rodriguez JJ, Whitefield JB, Hebart PDN (2006b) A minimalist barcode can identify a specimen whose DNA is degraded. Mol Ecol Notes 6:959–964. https://doi.org/10.1111/j.1471-8286.2006.01470.x
Hajibabaei M, Singer GAC, Hickey DA (2006c) Benchmarking DNA barcodes: an assessment using available primate sequences. Genome 49:851–854. https://doi.org/10.1139/G06.025
Hajibabaei M, Singer GAC, Clare EL, Hebert PDN (2007a) Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. BMC Biol 5:24. https://doi.org/10.1186/1741-7007-5-24
Hajibabaei M, Singer GAC, Hebert PDN, Hickey DA (2007b) DNA barcoding: how it complements taxonomy, molecular phylogenetic and population genetics. Trends Genet 23 (4):167–172. https://doi.org/10.1016/j.tig.2007.02.001
Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ (2011) Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS One 6(4):e17497. https://doi.org/10.1371/journal.pone.0017497
Halanych KM, Janosik AM (2006) A review of molecular markers used for Annelid phylogenetics. Integr Comp Biol 46:533–543
Hebert PDN, Barrett RDH (2005) Identifying spiders through DNA barcodes. Can J Zool 83:481–491
Hebert PDN, Gregory TR (2005) The promise of DNA barcoding for taxonomy. Syst Biol 54:852–859. https://doi.org/10.1080/10635150500354886
Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003a) Biological identifications through DNA barcodes. Proc R Soc Lond Ser B 270:313–321
Hebert PDN, Ratnasingham S, Ward JR (2003b) Barcoding animal life: cytochrome c oxidase subunit1 divergences among closely related species. Proc R Soc B 270:96–99
Hebert PDN, Penton EH, Burns JDH, Hallwachs W (2004a) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptesfulgreator*. Proc Natl Acad Sci U S A 101(41):14812–14817
Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004b) Identification of birds through DNA barcodes. PLoS Biol 2(1):1657–1663
Hebert PDNA, Alina C, Shelley Ball L, deWaard JR (2010) Biological identifications through DNA barcodes Biological identifications through DNA barcodes. Proc R Soc Lond 270:313–321
Hogg ID, Hebert PDN (2004) Biological identification of springtails (Collembola: Hexapoda) from the Canadian Arctic, using mitochondrial DNA barcodes. Can J Zool 82:1–6
Holmes BH, Steinke D, Ward RD (2009) Identification of shark and ray fins using DNA barcoding. Fish Res 95:280–288
Hopkins GW, Freckleton RP (2002) Declines in the numbers of amateur and professional taxonomists: implications for conservation. Anim Conserv 5:245–249
Hubert N, Hanner R, Holm E, Mandrak NE, Taylor E (2008) Identifying Canadian freshwater fishes through DNA barcodes. PLoS One 3(6):e2490. https://doi.org/10.1371/journal.pone.0002490
International Barcode of Life (2010) Lepidoptera barcode of life (Cited 15 Oct 2010). http://www.lepbarcoding.org/

Jarnegren J, Schander C, Sneli JA, Ronningen V, Young CM (2007) Four genes, morphology and ecology: distinguishing a new species of Acesta (Mollusca; Bivalvia) from the Gulf of Mexico. Mar Biol 152:43–55

Jayasankar P, Thomas PC, Paulton MP, Mathew J (2004) Morphometric and genetic analyzes of Indian mackerel (*Rastrelliger kanagurta*) from peninsular India. Asian Fish Sci 17:201–215

Jhon A, Prasanaakumar C, Lyla PS, Khan A, Jalal KCA (2010) DNA barcoding of *Lates calcarifer* (Bloch, 1970). Res J Biol Sci 5(6):414–419

Johnson GD, Keener P (1984) Aid to identification of American grouper larvae. Bull Mar Sci 34:106–134

Kartavtsev YP, Sharina SN, Goto T, Rutenko OA, Zemnukhov VV, Semenchenko AA, Pitruk DL, Hanzawa N (2009) Molecular phylogenetics of pricklebacks and other percoid fishes from the Sea of Japan. Aquat Biol 8:95–103. https://doi.org/10.3354/ab00205

Khedkar GD, Lutzky S, Kalyankar VB, Reddy ACS, David L (2009) DNA barcoding reveals a discontinuous genetic diversity pattern of fish in the Godavari River, India. In: 3rd international barcode of life conference, pp 80–81

Kocher TD, Thomas WK, Meyer A, Edwards SV, Pabo S, Villablanca FX, Wilson AC (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc Natl Acad Sci U S A 86:6196–6200

Kochzius M, Seidel C, Antoniou A, Botla SK, Campo D (2010) Identifying fishes through DNA barcodes and microarrays. PLoS One 5(9):e12620. https://doi.org/10.1371/journal.pone.0012620

Koedprang W, Nakorn UN, Nakajima M, Taniguchi H (2007) Evaluation of genetic diversity of eight grouper species *Epinephelus* spp. based on microsatellite variations. Fish Sci 73:227–236

Kumar CP, John BA, Khan SA, Lyla PS, Murugan S, Rozihan M, Jalal KCA (2011) Efficiency of universal barcode gene (Coxi) on morphologically cryptic mugilidae fishes delineation. Trends Appl Sci Res 6(9):1028–1036. https://doi.org/10.3923/tasr.2011.1028.1036

Lakra WS, Goswami M, Mohindra V, Lal KK, Punia P (2007) Molecular identification of five Indian Sciaenids (Pisces: perciformes, Sciaenidae) using RAPD markers. Hydrobiologia 583:359–363. https://doi.org/10.1007/s10750-006-0480-x

Lakra WS, Verma MS, Goswami M, Lal KK, Mohindra V, Punia P, Gopalakrishnan A, Singh KV, Ward RD, Hebert PDN (2010) DNA barcoding Indian marine fishes. Mol Ecol Resour 11:60–71

Lambert DM, Baker A, Huynen L, Haddrath O, Hebert PDN, Millar CD (2005) Is a large-scale DNA-based inventory of ancient life possible. J Hered 96(3):279–284

Langhoff P, Authier A, Buckley TR, Dugdale JS, Rodrigo A, Newcomb RD (2009) DNA barcoding of the endemic New Zealand leafroller moth genera, *Ctenopseustis* and *Planotortrix*. Mol Ecol Resour 9:691–698

Leasi F, Todaro MA (2009) Meiofaunal cryptic species revealed by confocal microscopy: the case of Xenotrichula intermedia (Gastrotricha). Mar Biol 156:1335–1346

Leis JM (1986) Larval development in four species of Indo-Pacific coral trout *Plectropomus* (Pisces: Serranidae: Epinephelinae) with an analysis of the relationships of the genus. Bull Mar Sci 38:525–552

Leis JM (2005) A larva of the elelinelutjanid, *Randallichythysfilamentosus* (Pisece: Perciformes), with comments on phylogenetic implications of larval morphology of basal lutjanids. Zootaxa 1008:57–64

Levin DM, Quinones ML, Povona MM, Linton Y (2006) DNA barcoding reveals both known and novel taxa in the Albitarsis Group (*Anopheles: Nyssorhynchus*) of Neotropical malaria vectors. Parasit Vectors. https://doi.org/10.1186/1756-3305-5-44

Lin CP, Danforth BN (2004) How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined data sets. Mol Phylogenet Evol 30 (3):686–702

Lorz AN, Linse K, Smith PJ, Steinke D (2012) First molecular evidence for underestimated biodiversity of Rhachotropis (Crustacea, Amphipoda), with description of a new species. PLoS One 7(3):e32365. https://doi.org/10.1371/journal.pone.0032365

Maggio T, Andaloro F, Hemida F, Arculeo M (2005) A molecular analysis of some Eastern Atlantic grouper from the *Epinephelus* and *Mycteroperca* genus. J Exp Mar Biol Ecol 321:83–92

Mecklenburg CW, Moller PR, Steinke D (2011) Biodiversity of arctic marine fishes: taxonomy and zoogeography. Mar Biodivers 41(1):109–140

Meusnier I, Singer ACG, Landry JF, Hickey AD, Hebert PDN, Hajibabaei M (2008) A universal DNA mini-barcoding for biodiversity analysis. BMC Genomics 9:214–218. https://doi.org/10.1186/147-2164-9-214

Meyer M, Briggs AW, Maricic T, Hober B, Hoffner B, Krause J, Weihmann A, Paabo S, Hofreiter M (2008) From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing. Nucleic Acids Res 36:e5

Miller TL, Cribb TH (2007) Phylogenetic relationship of some common Indo-Pacific snapper (Perciformes: Lutjanidae) based on mitochondrial DNA sequences, with comments on the taxonomic position of the Caesioninae. Mol Phylogenet Evol 44:450–460

Min XJ, Hickey DA (2007a) Assessing the effect of varying sequence length on DNA barcoding of fungi. Mol Ecol Notes. doi:https://doi.org/10.1111/j.1471-8286.2007.01698.x

Min XJ, Hickey DA (2007b) DNA barcodes provide a quick preview of mitochondrial genome composition. PLoS One 2(3):e325. https://doi.org/10.1371/journal.pone.0000325

Mitchell A (2008) DNA barcoding demystified. Aust J Entomol 47:169–173

Monaghan MT, Balke M, Gregory TR, Volger AP (2005) DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers. Philos Trans R Soc Lond B Biol Sci 360(1462):1925–1933. https://doi.org/10.1098/rstb.2005.1724. pp. 1–9

Moura T, Silva MC, Flgueiredo I, Neves A, Munoz PD, Coelho MM, Gordo LS (2008) Molecular bar-coding of north-east Atlantic deep-water sharks: species identification and application to fisheries management and conservation. Mar Freshw Res 59(3):214–223

Nelson JS (2006) Fishes of the world, 4th edn. Wiley, Hoboken, NJ

Nguyen HDT, Seifert KA (2008) Description and DNA barcoding of three new species of *Leohumicola* from South Africa and the United States. Persoonia 21:57–69. https://doi.org/10.3767/003158508X361334

Novacek M, Cleland EE (2001) The current biodiversity extinction event: scenarios for mitigation and recovery. Proc Natl Acad Sci U S A 98:5466–5470

Nylander JAA, Erséus C, Källersjo M (1999) A test of monophyly of the gutless Phallodrilinae (Oligochaeta, Tubificidae) and the use of a 573-bp region of the mitochondrial cytochrome oxidase I gene in analysis of annelid phylogeny. Zool Scr 28(3–4):305–313

Odeny DO, Oyieke HA, Ojwang WO (2009) DNA barcoding: refining para taxonomy for fishery surveys. 3rd international barcode of life conference, pp 81–82

Oliveira C, Pereira LHG, Henriques JM, Foresti F (2009) DNA barcode of freshwater fishes from upper Parana Basin, Brazil. 3rd international barcode of life conference, Mexico

Packer L, Gibbs J, Sheffield C, Hanner R (2009) DNA barcoding and the mediocrity of morphology. Mol Ecol Res 9:42–50

Pages N, Muñoz-Muñoz F, Talavera S, Sarto V, Lorca C, Núñez JI (2009) Identification of cryptic species of *Culicoides* (Diptera: Ceratopogonidae) in the subgenus *Culicoides* and development of species-specific PCR assays based on barcode regions. Vet Parasitol 165:298–310

Pegg CG, Sinclair BA, Briskey L, Aspden WJ (2006) MtDNA barcode identification of fish larvae in the southern great barrier reef, Australia. Sci Mar 70:7–12

Penton EH, Hebert PDN, Crease TJ (2004) Mitochondria, DNA variation in North American populations of Daphnia obtusa: continentalism or cryptic endemism? Mol Ecol 13(1):97–107

Persis M, Reddy ACS, Rao LM, Khedkar GD, Ravinder K, Nasruddin K (2009) COI (cytochrome oxidase – I) sequence based studies of Carangid fishes from Kakinada coast, India. Mol Biol Rep 36:1733–1740. https://doi.org/10.1007/s11033-008-9375-4

Pfenninger M, Cordellier M, Streit B (2006) Comparing the efficacy of morphologic and DNA-based taxonomy in the freshwater gastropod genus Radix (Basommatophora, Pulmonata). BMC Evol Biol 6:100

Pramual P, Wongpakam K, Adler PH (2011) Cryptic biodiversity and phylogenetic relationships revealed by DNA barcoding of Oriental black flies in the subgenus *Gomphostilbia* (Diptera: Simuliidae). Genome 54:1–9. https://doi.org/10.1139/G10-100

Radulovici AE, Archambault P, Dufresne F (2010) Review DNA barcodes for marine biodiversity: moving fast forward. Diversity 2:450–472. https://doi.org/10.3390/d2040450

Rajan PT (2010) Diversity of butterfly fishes (Chaetodontidae) of Andaman and Nicobar Islands: indicators in coral reef habitat monitoring and managements. In: Ramakrishna, Ragunanathan C, Sivaperuman C (eds) Recent trends in biodiversity of Andaman, Nicobar Islands. Zoological Survey of India, Kolkata, pp 337–342

Rajan PT, Sreeraj CR (2012) Structure of Reef fish communities of Seven Islands of Andaman and Nicobar Islands, India. In: Venkataraman K, Raghunathan C, Sivaperuman C (eds) Ecology of Faunal communities on the Andaman, Nicobar Islands. Springer, Berlin, pp 127–146

Ramakrishna, Raghunathan C, Sivaperuman C (2010) Biodiversity of Andaman and Nicobar Islands – an overview. In: Ramakrishna, Ragunanathan C, Sivaperuman C (eds) Recent trends in biodiversity of Andaman, Nicobar Islands. Zoological Survey of India, Kolkata, pp 1–42

Rao DV (2003) Guide to reef fishes of Andaman and Nicobar Islands. Zoological Survey of India, Kolkata, p 555

Rasmussen RS, Morrissey MT, Hebert PDN (2009) DNA barcoding of commercially important salmon and trout species in North America. In: 3rd international barcode of life conference, Mexico City, pp 84–85

Ratnasingham S, Hebert PDN (2007) BOLD: the barcode of life data system (www.barcodinglife.org). Mol Ecol Notes. doi:https://doi.org/10.1111/j.1471-8286.2006.01678.x

Remigio EA, Hebert PDN (2003) Testing the utility of partial COI sequences for phylogenetic estimates of gastropod relationships. Mol Phylogenet Evol 29(3):641–647

Richardson DE, Vanwye JD, Exum AM, Cowen RT, Crawford DL (2006) High-throughput species identification: from DNA isolation to bioinformatics. doi:https://doi.org/10.1111/j.1471-8286.2006.01620.x

Rivas LR (1949) A record of the Lutjanid fish (*Lutjanus cyanopterus*) for the Atlantic coast of the United States, with notes on related species of the genus. Copeia (2):150–152

Rock J, Costa FO, Walker DI, North AW, Hutchinson WF, Carvalho GR (2008) DNA barcodes of fish of the Scotia Sea, Antarctica indicate priority groups for taxonomic and systematic focus. Antarct Sci 20(3):253–262

Rubinoff PD (2006) Utility of mitochondrial DNA barcodes in species conservation. Conserv Biol 20(4):1026–1033

Saccone C, Giorgi CD, Gissi C, Pesole G, Reyes A (1999) Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. Gene 238:195–209

Sachithanandam V, Mohan PM, Dhivya P, Muruganandam N, Baskaran R, Chaaithanya IK, Vijayachari P (2011) DNA barcoding, phylogenetic relationships and speciation of Genus: *Plectropomus* in Andaman coast. J Res Biol 3:179–183

Sachithanandam V, Mohan PM, Muruganandam N, Chaaithanya IK, Dhivya P, Baskaran R (2012) DNA barcoding, phylogenetic study of *Epinephelus* spp from Andaman coastal region. Indian J Mar Sci 41(2):203–211

Salokannel J, Rantala M, Wahlberg N (2010) DNA-barcoding clarifies species definitions of Finnish *Apatania* (Trichoptera: Apataniidae). Entomol Fenn 2:1–11

Satheeshkumar P, Jagadeesan L (2010) Phylogenetic position and genetic diversity of Neridae – polychaeta based on molecular data from 16S rRNA sequences. Middle East J Sci Res 6 (6):550–555

Saunders G (2005) Applying DNA barcoding to red macroalgae, a preliminary appraisal holds promise for future applications. Philos Trans R Soc Ser Biol 360:1879–1888

Schander C, Willassen E (2005) What can biological barcoding do for marine biology. Mar Biol Res 1:79–83

Seifert KA, Samson RA, Dewaard JR, Houbraken J, Levesque CA, Moncalvo JM, Seize GL, Hebert PDN (2007) Prospects for fungus identification using CO1 DNA barcodes, with Penicillium as a test case. Proc Natl Acad Sci U S A 104(10):3901–3906

Shirak A, Zinder MC, Barroso RM, Seroussi E, Ron M, Hulata G (2009) DNA barcoding of Israeli Indigenous and introduced cichlids. Isr J Aquacult Bamidgeh 61(2):83–88

Siddall ME, Budinoff RB (2005) DNA-barcoding evidence for widespread introductions of a leech from South American *Helobdella triserialis* complex. Conserv Genet 6:467–472
Silva-Brandao KL, Lyra ML, Freitas AVL (2009) Barcoding Lepidoptera: current situation and perspectives on the usefulness of a contentious technique. Neotrop Entomol 38:441–451
Simon C et al (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann Entomol Soc Am 87(6):651–701
Simon C et al (2006) Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Annu Rev Ecol Evol Syst 37:545–579
Sjolin E, Erseus C, Kallersjo M (2005) Phylogeny of TubiWcidae (Annelida: Clitellata) based on mitochondrial and nuclear sequence data. Mol Phylogenet Evol 35:431–441
Smith WL, Craig MT (2007) Casting the percomorph net widely: the importance of broad taxonomic sampling in the search for the placement of serranid and percid fishes. Copeia 1:35–55
Smith PJ, Mcveagh SM, Steinke D (2008) DNA barcoding for the identification of smoked fish products. J Fish Biol 72(2):464–471
Steinke D, Zemlak TS, Boutillier JA, Hebert PDN (2009a) DNA barcoding of Pacific Canada's fishes. Mar Biol 156:2641–2647
Steinke D, Zemlak TS, Hebert PDN (2009b) Barcoding nemo: DNA-based identifications for the ornamental fish trade. PLoS One 4:e6300
Stoeckle MY, Hebert PDN (2008) Bar code of life: DNA tags help classify animals. Sci Am 299 (4):66–71
Sweeney BW, Battle JM, Jackson JK, Dapkey T (2011) Can DNA barcodes of stream macro invertebrates improve descriptions of community structure and water quality. J N Am Benthol Soc 30:195–216
Talwar PK (1990) Fishes of the Andaman and Nicobar Islands: a synoptic analysis. J Andaman Sci Assoc 6(2):71–102
Thomas M, Raharivololoniaina L, Glaw F, Vences M, Vieites DR (2005) Montane tadpoles in Madagascar: molecular identification and description of the larval stages of *Mantidactylus elegans*, *Mantidactylus madecassus*, and *Boophis laurenti* from the Andringitra Massif. Copeia 2005:174–183
Turanov SV, Kartavtseva IF, Zemnukhov VV (2012) Molecular phylogenetic study of several eelpout fishes (Perciformes, Zoarcoidei) from far eastern seas on the basis of the nucleotide sequence of the mitochondrial cytochrome oxidase 1 gene (Co-1). Genetika 48(2):235–252
Vargas SM, Araujo FCF, Santos FR (2009) DNA barcoding of Brazillian Sea turtles (Testudines). Genet Mol Biol 32(3):608–612
Venkataraman K, Wafar M (2005) Coastal and marine biodiversity of India. Ind J Mar Sci 34 (1):57–75
Victor BC, Hanner R, Shivji M, Hyde J, Caldow C (2009) Identification of the larval and juvenile stages of the Cubera Snapper, *Lutjanus cyanopterus*, using DNA barcoding. Zootaxa 2215:24–36
Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. Philos Trans R Soc Ser Biol 360:1847–1857
Ward RD, Costa FO, Holmes BH, Steinke D (2008) DNA barcoding of shared fish species from the North Atlantic and Australasia: minimal divergence for most taxa, but Zeus faberand *Lepidopus caudatus* each probably constitute two species. Aquat Biol 3:71–78
Ward RD, Hanner R, Hebert PDN (2009) The campaign to DNA barcode all fishes. FISH-BOL. J Fish Biol 74:329–356
Webb KE, Barnes DKA, Clark MS, Bowden DA (2006) DNA barcoding: a molecular tool to identify Antarctic marine larvae. Deep-Sea Res 53(II):1053–1060
Wiens JJ, Servedio MR (2000) Species delimitation in systematics: inferring diagnostic differences between species. Proc R Soc Lond B Biol 267:631–636
Will KW, Rubinoff D (2004) Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics 20:47–55

Wong EHK, Hanner R (2008) DNA barcoding detects market substitution in North American seafood. Food Res Int 41:828–837

Zemlak TS, Ward RD, Connell AD, Holmes BH, Hebert PDN (2009) Barcoding vertebrates DNA barcoding reveals overlooked marine fishes. Mol Ecol Res 9(1):237–242. https://doi.org/10.1111/j.1755-0998.2009.02649.x

Zettler L, Amaral A, Gómez F, Zettler E, Brendan GK, Amils R, Mitchell LS (2002) Eukaryotic diversity in Spain's river of fire. Nature 417:137

Zhang J (2011) Species identification of marine fishes in China with DNA barcoding. Evid Based Complement Alternat Med 21. https://doi.org/10.1155/2011/978253

Zhang JB, Hanner R (2011) DNA barcoding is a useful tool for the identification of marine fishes from Japan. Biochem Syst Ecol 39:31–42

Applications of DNA Barcoding in Fisheries

A. Pavan-Kumar, A. K. Jaiswar, P. Gireesh-Babu, A. Chaudhari, and G. Krishna

Abstract Sustainable management of fish resources rely on accurate identification of species for proper stock assessment, recruitment and delineation. Molecular markers would complement morphological tools and are able to delineate species more accurately. Mitochondrial cytochrome *c* oxidase subunit I has been standardized as a barcode gene for discriminating fishes. DNA barcoding approach can be applied in different facets of fisheries, and some of the applications include species diversity documentation, ichthyoplankton identification, fish prey items, invasive species identification, seed identification, parasite and vector identification and authentication of fish products. Further, with the advent of next-generation sequencing technology, it is possible to identify the presence of invasive species in environmental DNA collected from water and soil. In culture fisheries, some of the fish larvae survival is low due to lack of knowledge on their prey items. DNA barcoding approach with NGS technology could be useful to identify fish gut content and adulterated mixed fish products.

Keywords Cytochrome *c* oxidase subunit I · Invasive species · Seafood mislabelling · Conservation · Aquaculture

1 Introduction

Fisheries including capture and culture is one of the important sources of quality food and animal protein and provides livelihood for millions of people around the globe. Globally, fish and fishery products represent 9% of the total agricultural exports with a total production of 167.2 million tonnes (FAO 2016). Sustainability of fisheries depends on optimum exploitation of resources with effective management measures. Formulation of management measures relies on accurate identification of species for proper stock assessment, recruitment and delineation. Molecular markers would complement morphological tools and are able to delineate species more accurately. Mitochondrial cytochrome *c* oxidase subunit I (COI) gene has been

A. Pavan-Kumar (✉) · A. K. Jaiswar · P. Gireesh-Babu · A. Chaudhari · G. Krishna
ICAR-Central Institute of Fisheries Education, Mumbai, Maharashtra, India

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_17

standardized as a barcode gene for discriminating metazoans (Hebert et al. 2003). After successful demonstration of efficiency of DNA barcoding in delineating fishes (Ward et al. 2005), its application has been percolated to different facets of fisheries.

2 DNA Barcoding Applications in Capture Fisheries

2.1 Species Identification

Fishes, one of the largest vertebrate groups with more than 32,000 species, display a spectrum of morphological characters and have been reported to show phenotypic plasticity (Muschick et al. 2011; Barry et al. 2016). Accurate identification and cataloguing of fishes is necessary for sustainable fishery management. DNA barcoding has been primarily used to discriminate the fish species based on intra- and interspecific genetic divergence values. Reference barcodes have been developed for 17,108 fish species (50%), and several cryptic species have been described in different fish groups (Barcode of Life Data Systems 2017; Pavan-Kumar et al. 2016). DNA barcoding has also become one of the discernible characters for describing new species, and 32 new species have been described over the last 2 years (Table 1). DNA barcodes are also useful to resolve taxonomic ambiguity by investigating the evolutionary relationship among lower taxonomic rank (i.e. up to species level).

2.2 Ecosystem Management

The biodiversity of different sensitive ecosystems such as coral reefs, mangroves and wetlands are under threat due to anthropogenic factors, invasive alien species and climate change (Masters and Norgrove 2010). These factors cause habitat destruction, loss of biodiversity and subsequently reduction of resilience of ecosystem. Regular monitoring of biodiversity from these ecosystems would provide vital information about pattern of changes in biodiversity and impact of environment on native biodiversity. To achieve this, baseline information on native biodiversity is essential, and DNA barcoding/metabarcoding would be the best approach for developing ecosystem-specific databases. Leray and Knowlton (2015) used DNA barcoding and metabarcoding approach to characterize benthic diversity on oyster reefs.

2.3 Ichthyoplankton Identification

To understand the processes that influence species spatial distribution, population dynamics and migration strategies, it is essential to study larvae ecology and transport (Bakun 1996; Cowen et al. 2006). Ichthyoplankton collection and identification up to

Table 1 List of new species described along with DNA barcodes

S. No	Order	Family	Species	References
1	Perciformes	Labridae	*Paracheilinus paineorum*	Allen et al. (2016)
2			*Paracheilinus xanthocirritus*	
3			*Paracheilinus alfiani*	
4		Pempheridae	*Pempheris gasparinii*	Pinheiro et al. (2016)
5		Serranidae	*Tosanoides obama*	Pyle et al. (2016)
6		Serranidae	*Chelidoperca santosi*	Williams and Carpenter (2015)
7		Cichlidae	*Ptychochromis mainty*	Martinez et al. (2015)
8		Badidae	*Badis britzi*	Dahanukar et al. (2015)
9	Anguilliformes	Nettastomatidae	*Saurenchelys gigas*	Lin et al. (2015)
10		Muraenidae	*Gymnothorax pseudomelanosomatus*	Loh et al. (2015)
11	Aulopiformes	Synodontidae	*Trachinocephalus gauguini*	Guimarães-Costa et al. (2016)
12	Clupeiformes	Engraulidae	*Stolephorus tamilensis*	Gangan (2017)
13	Cypriniformes	Cyprinidae	*Squalius namak*	Khaefi et al. (2016)
14			*Danio htamanthinus*	Kullander et al. (2015)
15			*Garra mondica*	Sayyadzadeh et al. (2015)
16			*Coreoleuciscus aeruginos*	Song and Bang (2015)
17			*Metzia parva*	Luo et al. (2015)
18			*Danio annulosus*	Kullander et al. (2015)
19			*Osteobrama serrata*	Singh et al. (2017)
20			*Tor dongnaiensis*	Hoàng et al. (2015)
21			*Alburnoides damghan*	Roudbar et al. (2016)
22			*Capoeta coadi*	Alwan et al. (2016)
23			*Hypselobarbus bicolor*	Knight et al. (2016)
24		Cobitidae	*Cobitis avicennae*	Mousavi-Sabet et al. (2015)
25		Nemacheilidae	*Eidinemacheilus proudlovei*	Freyhof et al. (2016)
26	Rajiformes	Rajiidae	*Raja parva*	Last and Séret (2016)
27	Myliobatiformes	Dasyatidae	*Maculabatis ambigua*	Last et al. (2016)

(continued)

Table 1 (continued)

S. No	Order	Family	Species	References
28	Decapoda	Palaemonidae	*Hamodactylus paraqabai*	Horka et al. (2016)
29			*H. pseudaqabai*	Horka et al. (2016)
30		Alvinocarididae	*Alvinocaris kexueae*	Wang and Sha (2016)
31		Macrophthalmidae	*Macrophthalmus purpureocheir*	Teng and Shih (2015)
32	Pleurotomarioidea	Pleurotomariidae	*Bayerotrochus delicatus*	Zhang et al. (2016)

species level would help in locating spawning areas (Serafy et al. 2003; Govoni et al. 2003), population size estimation (Ralston et al. 2003) and understanding the distribution of rare and cryptic species (Richardson and Cowen 2004). Fish egg/larvae can be identified by comparing the COI sequences of eggs/larvae against reference sequence libraries such as BOLD (Barcode of Life Data Systems). The sequences will be assigned to taxon based on their similarity/genetic distance values against database sequences.

DNA barcoding approach has been successfully applied for identification of ichthyoplankton from Antarctica (Webb et al. 2006), Great Barrier Reef (Pegg et al. 2006), Yucaton Peninsula, Mexico (Valdez-Moreno et al. 2010) and coral reefs (Hubert et al. 2015). Several studies implied this approach for describing life history traits of new species (Victor 2007; Victor et al. 2009) and for identifying fish spawning areas (Neira et al. 2014). Some of the studies have compared the efficiency of DNA barcoding approach against visual identification techniques and reported superior performance of DNA barcoding in ichthyoplankton identification (Ko et al. 2013; Becker et al. 2015; Puncher et al. 2015; Overdyk et al. 2016).

2.4 Conservation

DNA barcoding is a powerful tool for rapid and accurate species identification in an economic and effective way. DNA barcodes are very useful for characterizing biodiversity in ecosystems that are species-rich, difficult to assess and poorly catalogued. DNA barcodes can be used to estimate phylogenetic diversity values. Phylogenetic diversity (PD) values are the minimum total length of all the phylogenetic branches required to span a given set of taxa on the phylogenetic tree (Faith 1992). Faith and Baker (2007) advocated consideration of PD values for prioritization of geographical locations for conservation purpose. Those areas that contribute much to the PD values should be given preference for conservation purpose.

As per IUCN guidelines, criteria for assessing the conservation status of any species are its population size, extent of occurrence, area of occupancy, etc. Along with other factors, DNA barcoding would help in resolving taxonomy and assist in

prioritizing species for conservation purpose. For example, consider two species "species A" and "species B" that are showing similar morphological characters and have restricted distribution. If they are assessed based on morphological characters only, these two species could be considered as a single species and the conservation status be assigned as "least concern". In this case, implementation of DNA barcoding can resolve the taxonomic ambiguity, and species would be properly evaluated for conservation status.

2.5 *Mislabelling of Seafood Products*

Nutrient profile of each fish is different, and the price should be at par with nutrient values and should be labelled properly (Mohanty et al. 2014; Bogard et al. 2015). However, as with other processed products, mislabelling, adulteration and replacement of high-value fish with low-value fish have been reported in seafood processed products (Armani et al. 2011; Marko et al. 2004). Since the processed products lack morphological characters, it is impossible to identify the species. DNA barcoding has been extensively used to authenticate seafood processed products and reported various levels of mislabelling globally (Nagalakshmi et al. 2016; Cutarelli et al. 2014; Harris et al. 2016). The level of DNA degradation would be relatively high in thermal processed products, and amplification of the entire barcode region (650 b) is relatively difficult. In this scenario, mini barcodes (250–300 bp) have been developed and successfully used for identification of thermal processed products (Shokralla et al. 2015). In case of mixed samples, DNA metabarcoding approach (integration of next-generation sequencing with DNA barcoding) can be employed.

2.6 *Biosecurity and Invasive Species*

Biosecurity (prevention of the entry and spread of infectious diseases, genetically modified organisms and biological weapons) measures have to be implemented not only for disease-causing agents but also for invasive alien species which cause huge economic and ecosystem loss (Williamson 1996). The risk of invasion by non-native species is increased by international trade, and invasive species colonization is influenced by land use and climate change (Masters and Norgrove 2010). Several studies have reported the detrimental effects of invasive alien species on native biodiversity and subsequent economic loss (Chapin et al. 2000; Pimentel et al. 2005). Rapid and accurate identification of traded biological materials up to species level at the port of entry is one of the important factors for implementing biosecurity measures. DNA barcodes would assist in precise identification of species, and reference barcodes have been developed for several commercially important fish and shellfish (Bamaniya et al. 2016; Collins et al. 2012; BOLD 2017). Marescaux and Van Doninck (2013) discriminated two invasive species *Dreissena polymorpha*

and *D. rostriformis bugensis* using DNA barcodes. Metabarcoding of environmental DNA (eDNA) has been successfully used to detect invasive species from different water bodies (Frédéric et al. 2015; Furlan et al. 2015; Salisbury et al. 2015).

2.7 Predator-Prey Relationships

Predator-prey interactions play an important role in fisheries management. As per the Pope's study (1979), the overall maximum sustainable yield (MSY) for the system will be higher if the fish species in an ecosystem are linked mainly through predator-prey relationships than species are competing with each other. For many fish species, predation during early life stages has been reported to be one of the main factors in restraining recruitment success (Saitoh et al. 2003). Further, predator-prey interactions provide vital information on habitat use and identify critical foraging habitats (Peters et al. 2014). Size class-wise information on prey items would be useful to minimize predator impact on the native fish populations they prey on. Jo et al. (2014) identified prey items of largemouth bass (*Micropterus salmoides*) in different size classes using DNA barcoding approach and showed variation in prey preference as they mature (ontogenetic diet shift).

Inadvertent release of non-native species (invasive species) could cause restructuring of ecosystem by changing food webs and altering the pattern of resource utilization (Stigall 2012). Most often invasive species are generalists and are flexible with the prey items (Olden et al. 2004). Côté et al. (2013) characterized prey items of invasive Indo-pacific lionfish (*Pterois volitans*) from Bahamian coral reefs through DNA barcoding approach and reported 37 fish prey species. Moran et al. (2016) characterized prey items of non-native blue catfish (*Ictalurus furcatus*) and flathead catfish (*Pylodictis olivaris*) and showed the presence of economically and ecologically important species in their diet. Metabarcoding approach, an extension of DNA barcoding method wherein PCR amplicons from mixed/bulk samples are sequenced using next-generation sequencing platform, has been successfully applied to characterize prey items from different fishes (Leray et al. 2013; Harms-Tuohy et al. 2016). Sousa et al. (2016) analysed gut content of ocean sun fish *Mola mola* (world's heaviest bony fish) through metabarcoding approach. This study proves that *Mola mola* is a generalist predator and identified 41 different prey items.

3 Potential Applications of DNA Barcoding in Aquaculture

3.1 Seed Identification

Accurate seed identification could be possible by metabarcoding of DNA collected from hatchery water.

3.2 Fingerling Survival

In aquaculture, for some of the fishes, larvae or fingerling survival is low, and one of the reasons for this is lack of proper nutrition/feed. Providing natural feed which the species get from the wild would be a promising approach for better survival. However, for most of the culture species, the information of prey items during different ontogenic stages is lacking. DNA metabarcoding of gut content of fingerling/adult/brooder fishes would give insights about their natural prey items. The prey organisms once identified can be cultured and provided as live feed to fish fingerlings.

3.3 Improved Management/Production Management

Biofloc technology (BFT) is going to revolutionize aquaculture industry by promoting waste retention and its subsequent conversion to biofloc as a natural food for fish/shrimps (Crab et al. 2012). Biofloc is a consortium of microorganism such as heterotrophic bacteria, algae (dinoflagellates and diatoms), fungi, ciliates, flagellates, rotifers, nematodes and metazoans. They act synergistically to maintain the water quality and convert nitrogenous waste into protein. The species composition of biofloc varies with carbon source, and characterization of biofloc microscopic composition would be very useful for better management of culture ponds. Metabarcoding of biofloc could reveal the composition of biofloc, and this information along with carbon source would act as reference for implementing biofloc technology effectively. Different genes, i.e. internal transcribed spacer (ITS), COI, 16S rRNA and LSU D1/D2 have been standardized for developing barcodes for fungi, protozoans/metazoans and bacteria, respectively (Lebonah et al. 2014; Schoch et al. 2012).

3.4 Health Management

Pond ecosystem is dynamic, and any alteration in biotic and abiotic factors could result in stress on fishes/shrimps and may increase the susceptibility of fish to infections. Metabarcoding of soil and water at regular intervals would give information about changes in bacteria/microbe composition over the time with response to uneaten, leftover feed and other activities. This information would act as a reference for better management of ponds by keeping optimum water and soil quality parameters. Parasites (monogeneans, digeneans and crustaceans) infection in fishes causes huge economic losses due to secondary bacterial infection and subsequent mortality. These parasites can be eradicated if early detection is possible. However, larvae and eggs of these parasites are very small in size and cannot be

identified by the naked eye. Metabarcoding of DNA from pond water can be useful to identify early life stages of fish parasites. Further, generated DNA barcodes are useful to study the evolution of host-parasite interaction. Several authors have generated DNA barcodes for monogeneans (Hansen et al. 2007; Neeraja et al. 2016), digeneans (Locke et al. 2015) and crustaceans (Ferozkhan et al. 2016).

4 Conclusion

Despite few limitations (lack of DNA barcoding gap in recently evolved species, unable to identify hybrid species), DNA barcoding approach is the most promising approach to document the biodiversity and play major role in achieving Convention on Biodiversity aichi targets. Implementation of DNA barcoding approach in different fields of fisheries would assist in sustainable utilization of resources and increased culture production.

References

Allen GR, Erdmann MV, Yusmalinda NLA (2016) Review of the Indo-Pacific Flasher wrasses of the genus Paracheilinus (Perciformes: Labridae), with descriptions of three new species. J Ocean Sci Found 19:18–90

Alwan NH, Zareian H, Esmaeili HR (2016) *Capoeta coadi*, a new species of cyprinid fish from the Karun River drainage, Iran based on morphological and molecular evidences (Teleostei, Cyprinidae). Zookeys 16:155–180

Armani A, Castigliego L, Tinacci L, Gianfaldoni D, Guidi A (2011) Molecular characterization of icefish (Salangidae family), using direct sequencing of mitochondrial cytochrome b gene. Food Control 22:888–895

Bakun A (1996) Patterns in the ocean: ocean processes and marine population dynamics. California sea grant college system report T-o37, La Jolla, CA

Bamaniya DC, Pavan-Kumar A, Gireesh-Babu P, Sharma N, Reang D, Krishna G, Lakra WS (2016) DNA barcoding of marine ornamental fishes from India. Mitochondrial DNA A DNA Mapp Seq Anal 27:3093–3097

Barcode of Life Database (2017) http://boldsystems.org/. Accessed 10 Mar 2017

Barry J, Newton M, Dodd JA, Hooker OE, Boylan P, Lucas MC, Adams CE (2016) Foraging specialisms influence space use and movement patterns of the European eel *Anguilla anguilla*. Hydrobiologia 766:333–348

Becker RA, Sales NG, Santos GM, Santos GB, Carvalho DC (2015) DNA barcoding and morphological identification of neotropical ichthyoplankton from the Upper Paraná and São Francisco. J Fish Biol 87:159–168

Bogard JR, Thilsted SH, Marks GC, Wahab MA, Hossain MAR, Jakobsen J, Stangoulis J (2015) Nutrient composition of important fish species in Bangladesh and potential contribution to recommended nutrient intakes. J Food Compos Anal 42:120–133

Chapin FS III, Zavaleta ES, Eviner VT, Naylor RL, Vitousek PM et al (2000) Consequences of changing biodiversity. Nature 405:234–242

Collins RA, Armstrong KF, Meier R, Yi Y, Brown SDJ et al (2012) Barcoding and border biosecurity: Identifying cyprinid fishes in the aquarium trade. PLoS One 7(1):e28381. https://doi.org/10.1371/journal.pone.0028381

Côté IM, Green SJ, Morris JA Jr, Akins JL, Steinke D (2013) Diet richness of invasive Indo-Pacific lionfish revealed by DNA barcoding. Mar Ecol Prog Ser 472:249–256
Cowen RK, Paris CB, Srinivasan A (2006) Scaling connectivity in marine populations. Science 311:522–527
Crab R, Defoirdt T, Bossier P, Verstraete W (2012) Biofloc technology in aquaculture: beneficial effects and future challenges. Aquaculture 356–357:351–356
Cutarelli A, Amoroso MG, De Roma A, Girardi S, Galiero G, Guarino A et al (2014) Italian market fish species identification and commercial frauds revealing by DNA sequencing. Food Control 37:46–50
Dahanukar N, Kumkar P, Katwate U, Raghavan R (2015) *Badis britzi*, a new percomorph fish (Teleostei: Badidae) from the Western Ghats of India. Zootaxa 3941:429–436
Faith DP (1992) Conservation evaluation and phylogenetic diversity. Biol Conserv 61:1–10
Faith DP, Baker AM (2007) Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. Evol Bioinformatics Online 2:121–128
FAO (2016) The state of world fisheries and aquaculture 2016. Contributing to food security and nutrition for all. FAO, Rome, 200 p
Ferozkhan K, Sanker G, Prasanna Kumar C (2016) Linking eggs and adults of Argulus spp. using mitochondrial DNA barcodes. Mitochondrial DNA A DNA Mapp Seq Anal 27:3927–3931
Frédéric JJ, Chain, Brown E, MacIsaac H, Melania E (2015) Monitoring biodiversity for the early detection of aquatic invasive species using metabarcoding applied across Canadian ports in the Pacific, Arctic, Atlantic, and Great Lakes. In: Scientific abstracts from the 6th international barcode of life conference. Genome 58(5):245
Freyhof J, Abdullah YS, Ararat K, Ibrahim H, Geige MF (2016) Eidinemacheilus proudlovei, a new subterranean loach from Iraqi Kurdistan (Teleostei; Nemacheilidae). Zootaxa 4173:225–236
Furlan E, Gleeson D, Hardy C, Duncan R (2015) A framework for estimating eDNA sensitivity. In: Scientific abstracts from the 6th international barcode of life conference. Genome 58(5):218
Gangan S (2017) Taxonomic evaluation of species of family Engraulidae occurring in Indian waters PhD Thesis. ICAR-Central Institute of Fisheries Education, Mumbai
Govoni JJ, Laban EH, Hare JA (2003) The early life history of swordfish (*Xiphias gladius*) in the western North Atlantic. Fish Bull 101:778–789
Guimarães-Costa A, Vallinoto M, Giarrizzo T, Pezold F, Schneider H, Sampaio I (2016) Molecular evidence of two new species of Eleotris (Gobiiformes: Eleotridae) in the western Atlantic. Mol Phylogenet Evol 98:52–60
Hansen H, Bakke TA, Bachmann L (2007) Mitochondrial haplotype diversity of *Gyrodactylus thymalli* (Platyhelminthes; Monogenea): extended geographic sampling in United Kingdom, Poland, and Norway reveals further lineages. Parasitol Res 100:1389–1394
Harms-Tuohy AC, Schizas VN, Appeldoorn SR (2016) Use of DNA metabarcoding for stomach content analysis in the invasive lionfish *Pterois volitans* in Puerto Rico. Mar Ecol Prog Ser 558:181–191
Harris JD, Rosado D, Xavier R (2016) DNA barcoding reveals extensive mislabeling in seafood sold in Portuguese supermarkets. J Aquat Food Prod T 25:1375–1380
Hebert PDN, Cywinska A, Ball SL (2003) Biological identifications through DNA barcodes. Proc Biol Sci R Soc 270(1512):313–321. https://doi.org/10.1098/rspb.2002.2218
Hoàng HĐ, Phạm HM, Durand JD, Trần NT, Phan PĐ (2015) Mahseers genera Tor and Neolissochilus (Teleostei: Cyprinidae) from southern Vietnam. Zootaxa 4006:551–568
Horka I, Fransen MJH, Duris Z (2016) Two new species of shrimp of the Indo-West Pacific genus *Hamodactylus Holthuis*, 1952 (Crustacea: Decapoda: Palaemonidae). Eur J Taxon 188:1–26
Hubert N, Espiau B, Meyer C Planes S (2015) Identifying the ichthyoplankton of a coral reef using DNA barcodes. Mol Ecol Resour 15:57–67
Jo H, Gim JA, Jeong KS, Kim HS, Joo GJ (2014) Application of DNA barcoding for identification of freshwater carnivorous fish diets: Is number of prey items dependent on size class for Micropterus salmoides? Ecol Evol 4:219–229

Khaefi R, Esmaeili HR, Sayyadzadeh G, Geiger MF, Freyhof J (2016) *Squalius namak*, a new chub from Lake Namak basin in Iran (Teleostei: Cyprinidae). Zootaxa 4169:145–159

Knight JD, Rai A, D'Souza RK, Philip S, Dahanukar N (2016) *Hypselobarbus bicolor*, a new species of large barb (Teleostei: Cyprinidae) from the Western Ghats of India. Zootaxa 4184:316–328

Ko HL, Wang YT, Chiu TS, Lee MA, Leu MY, Chang KZ, Chen WY, Shao KT (2013) Evaluating the accuracy of morphological identification of larval fishes by applying DNA barcoding. PLoS One 8:e53451

Kullander SO, Rahman MM, Norén M, Mollah AR (2015) *Danio annulosus*, a new species of chain Danio from the Shuvolong Falls in Bangladesh (Teleostei: Cyprinidae: Danioninae). Zootaxa 3994:053–068

Last PR, Séret B (2016) A new Eastern Central Atlantic skate *Raja parva* sp. nov. (Rajoidei: Rajidae) belonging to the *Raja miraletus* species complex. Zootaxa 4147:477–489

Last PR, Bogorodsky SV, Alpermann TJ (2016) *Maculabatis ambigua* sp. nov., a new whipray (Myliobatiformes: Dasyatidae) from the Western Indian Ocean. Zootaxa 4154:66–78

Lebonah DE, Dileep A, Chandrasekhar K, Sreevani S, Sreedevi B, Kumari JP (2014) DNA barcoding on bacteria: a review. Adv Biol 2014:1–9

Leray M, Knowlton N (2015) DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. Proc Natl Acad Sci U S A 112:2076–2081

Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N et al (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Front Zool 10:34

Lin J, Smith DG, Shao KT, Chen HM (2015) *Saurenchelys gigas* sp. nov., a new nettastomatid eel (Teleostei, Anguilliformes, Nettastomatidae) from the western central Pacific. Zootaxa 4060:121–130

Locke SA, Al-Nasiri FS, Caffara M, Drago F, Kalbe M, Lapierre AR, McLaughlin JD (2015) Diversity, specificity and speciation in larval Diplostomidae (Platyhelminthes: Digenea) in the eyes of freshwater fish, as revealed by DNA barcodes. Int J Parasitol 45:841–855

Loh KH, Shao KT, Ho HC, Lim PE, Chen HM (2015) A new species of moray eel (Anguilliformes: Muraenidae) from Taiwan, with comments on related elongate unpatterned species. Zootaxa 4060:30–40

Luo W, Sullivan JP, Zhao HT, Peng ZG (2015) *Metzia parva*, a new cyprinid species (Teleostei: Cypriniformes) from south China. Zootaxa 3962:226–234

Marescaux J, Van Doninck K (2013) Using DNA barcoding to differentiate invasive Dreissena species (Mollusca, Bivalvia). In: Nagy ZT, Backeljau T, De Meyer M, Jordaens K (Eds) DNA barcoding: a practical tool for fundamental and applied biodiversity research. ZooKeys 365: 235–244

Marko PB, Lee SC, Rice AM, Gramling JM, Fitzhenry TM, McAlister JS, Harper GR, Moran AL (2004) Fisheries: mislabelling of a depleted reef fish. Nature 430:309–310

Martinez CM, Arroyave J, Sparks JS (2015) A new species of Ptychochromis from southeastern Madagascar (Teleostei: Cichlidae). Zootaxa 4044:79–92

Masters G, Norgrove L (2010) Climate change and invasive alien species. CABI working paper 1, 30 p

Mohanty B, Mahanty A, Ganguly S et al (2014) Amino acid compositions of 27 food fishes and their importance in clinical nutrition. J Amino Acids. https://doi.org/10.1155/2014/269797

Moran Z, Orth DJ, Schmitt JD, Hallerman EM, Aguilar R (2016) Effectiveness of DNA barcoding for identifying piscine prey items in stomach contents of piscivorous catfishes. Environ Biol Fish 99:161–167. https://doi.org/10.1007/s10641-015-0448-7

Mousavi-Sabet H, Vatandoust S, Esmaeili HR, Geiger MF, Freyhof J (2015) *Cobitis avicennae*, a new species of spined loach from the Tigris River drainage (Teleostei: Cobitidae). Zootaxa 3914:558–569

Muschick M, Barluenga M, Salzburger W, Meyer A (2011) Adaptive phenotypic plasticity in the Midas cichlid fish pharyngeal jaw and its relevance in adaptive radiation. BMC Evol Biol 11:116

Nagalakshmi K, Pavan-Kumar A, Venketashwarlu G, Gireesh-Babu P, Lakra WS (2016) Mislabeling in Indian seafood: an investigation using DNA barcoding. Food Control 59:196–200
Neeraja T, Tripathi G, Gireesh-Babu P, Pavan-Kumar A (2016) Molecular characterization and phylogeny of some mazocraeidean monogeneans from carangid fish. Acta Parasitol 61:360–368
Neira FJ, Perry RA, Burridge CP, Lyle JM, Keane JP (2014) Molecular discrimination of shelf-spawned eggs of two co-occurring *Trachurus* spp.(Carangidae) in southeastern Australia: a key step to future egg-based biomass estimates. ICES J Mar Sci 72:614–624
Olden JD, Poff NL, Douglas MR, Douglas ME, Fausch KD (2004) Ecological and evolutionary consequences of biotic homogenization. Trends Ecol Evol 19:18−24
Overdyk LM, Holm E, Crawford SS, Hanner RH (2016) Increased taxonomic resolution of Laurentian Great Lakes ichthyoplankton through DNA barcoding: a case study comparison against visual identification of larval fishes from Stokes Bay, Lake Huron. J Great Lakes Res 42:812–818
Pavan-Kumar A, Gireesh-Babu P, Jaiswar AK, Chaudhari A, Krishna G, Lakra WS (2016) DNA barcoding of marine fishes: prospects and challenges. In: Trivedi S, Ansari A, Ghosh S, Rehman H (eds) DNA barcoding in marine perspectives. Springer, Cham. https://doi.org/10.1007/978-3-319-41840-7_17
Pegg GG, Sinclair B, Briskey L, Aspden WJ (2006) MtDNA barcode identification of fish larvae in the southern Great Barrier Reef–Australia. Sci Mar 70:7–12
Peters KJ, Ophelkeller K, Bott NJ, Deagle BE, Jarman SN, Goldsworthy SD (2014) Fine-scale diet of the Australian sea lion (*Neophoca cinerea*) using DNA-based analysis of faeces. Mar Ecol 36:347–367. https://doi.org/10.1111/maec.12145
Pimentel D, Zuniga R, Morrison D (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. Ecol Econ 52:273–288
Pinheiro HT, Bernardi G, Rocha LA (2016) *Pempheris gasparinii*, a new species of sweeper fish from Trindade Island, southwestern Atlantic (Teleostei, Pempheridae). Zookeys 8:105–115
Pope JG (1979) Stock assessment in multispecies fisheries, with special reference to the trawl fishery in the Gulf of Thailand. SCS/DEV.79/19, South China Sea Fisheries Development and Coordinating Programme, 106 p
Puncher G, Alemany F, Arrizabalaga H, Cariani A, Tinti F (2015) Misidentification of bluefin tuna larvae: a call for caution and taxonomic reform. Rev Fish Biol Fish 25:485–502
Pyle LR, Green DB, Kosaki KR (2016) *Tosanoides obama*, a new basslet (Perciformes, Percoidei, Serranidae) from deep coral reefs in the Northwestern Hawaiian Islands. ZooKeys 641:165–181
Ralston S, Bence JR, Eldridge MB, Lenarz WH (2003) An approach to estimating rockfish biomass based on larval production, with application to *Sebastes jordani*. Fish Bull 101:129–146
Richardson DE, Cowen RK (2004) Diversity of leptocephalus larvae around the island of Barbados (West Indies): relevance to regional distributions. Mar Ecol Prog Ser 282:271–284
Roudbar AJ, Eagderi S, Esmaeili HR, Coad BW, Bogutskaya N (2016) A molecular approach to the genus Alburnoides using COI sequences data set and the description of a new species, *A. damghani*, from the Damghan River system (the Dasht-e Kavir Basin, Iran) (Actinopterygii, Cyprinidae). Zookeys 579:157–181
Saitoh K, Takagaki M, Yamashita Y (2003) Detection of Japanese flounder-specific DNA from gut contents of potential predators in the field. Fish Sci 69:473–477
Salisbury S, Shokralla S, Hajibabaei M (2015) The feasibility of detecting an Asian carp invasion using environmental DNA and next-generation sequencing. In: Scientific abstracts from the 6th international barcode of life conference. Genome 58(5):275
Sayyadzadeh G, Esmaeili HR, Freyhof J (2015) *Garra mondica*, a new species from the Mond River drainage with remarks on the genus Garra from the Persian Gulf basin in Iran (Teleostei: Cyprinidae). Zootaxa 4048:75–89
Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc Natl Acad Sci U S A 109:6241–6246
Serafy JE, Cowen RK, Paris CB, Capo TR, Luthy SA (2003) Evidence of blue marlin, *Makaira nigricans*, spawning in the vicinity of Exuma Sound, Bahamas. Mar Freshw Res 54:299–306

Shokralla S, Hellberg RS, Handy SM, King I, Hajibabaei MA (2015) DNA mini-barcoding system for authentication of processed fish products. Sci Rep 5:15894

Singh M, Verma R, Yumnam R, Vishwanath W (2017) Molecular phylogenetic analysis of genus Osteobrama Heckel, 1843 and discovery of *Osteobrama serrata* sp. nov. from North East India. Mitochondrial DNA A DNA Mapp Seq Anal. https://doi.org/10.1080/24701394.2017.1285289

Song HY, Bang IC (2015) *Coreoleuciscus aeruginos* (Teleostei: Cypriniformes: Cyprinidae), a new species from the Seomjin and Nakdong rivers, Korea. Zootaxa 3931:140–150

Sousa LL, Xavier R, Costa V, Humphries EN, Trueman C, Rosa R, Sims WD, Queiroz N (2016) DNA barcoding identifies a cosmopolitan diet in the ocean sunfish. Sci Rep 6:28762

Stigall LA (2012) Invasive species and evolution. Evol Educ Outreach 5:526–533

Teng SJ, Shih HT (2015) A new species of *Macrophthalmus desmarest*, 1823 (Crustacea: Decapoda: Brachyura: Macrophthalmidae) from Taiwan, with notes on four new records. Zootaxa 4058:451–470

Valdez-Moreno M, Vásquez-Yeomans L, Elías-Gutiérrez M, Ivanova NV, Hebert PDN (2010) Using DNA barcodes to connect adults and early life stages of marine fishes from the Yucatan Peninsula, Mexico: potential in fisheries management. Mar Freshw Res 61:655–671

Victor BC (2007) *Coryphopterus kuna*, a new goby (Perciformes: Gobiidae: Gobiinae) from the western Caribbean, with the identification of the late larval stage and an estimate of the pelagic larval duration. Zootaxa 1526:51–61

Victor BC, Hanne R, Shivji M, Hyde J, Caldow C (2009) Identification of the larval and juvenile stages of the cubera snapper, *Lutjanus cyanopterus,* using DNA barcoding. Zootaxa 2215:24–36

Wang YR, Sha ZL (2016) A new species of the genus Alvinocaris Williams and Chace, 1982 (Crustacea: Decapoda: Caridea: Alvinocarididae) from the Manus Basin hydrothermal vents, Southwest Pacific. Zootaxa. https://doi.org/10.11646/zootaxa.4226.1.7

Ward RD, Zemlak TS, Innes BH (2005) Barcoding Australia's fish species. Philos Trans R Soc Lond B 360:1847–1857

Webb KE, Barnes DK, Clark MS, Bowden DA (2006) DNA barcoding: a molecular tool to identify Antarctic marine larvae. Deep Sea Res Part 2 Top Stud Oceanogr 53:1053–1060

Williams TJ, Carpenter EK (2015) A new fish species of the subfamily Serraninae (Perciformes, Serranidae) from the Philippines. Zootaxa 3911:287–293

Williamson M (1996) Biological invasions. Chapman & Hall, London, 244 p

Zhang S, Zhang S, Wei P (2016) *Bayerotrochus delicatus*, a new species of pleurotomariid from Yap Seamount, near Palau, Western Pacific (Gastropoda: Pleurotomariidae). Zootaxa 4161:252–260

Identification and Conservation of Reptiles Through DNA Barcoding

Subrata Trivedi, Hasibur Rehman, Shalini Saggu, Al Thabiani Aziz, Chellasamy Panneerselvam, and Sankar K. Ghosh

Abstract Reptiles are cold-blooded vertebrates that play very important ecological role in the environment. They dominated the world in the Mesozoic era and many of them have become extinct, and some are facing extinction. Poaching, overexploitation, secondary poisoning, habitat loss, climate change, etc. are causing serious threats to the biodiversity of the reptiles. The taxonomic position of the reptiles also makes this group evolutionarily very significant. In this chapter, we discuss the role of DNA barcoding in the identification, phylogenetic assessment, and conservation of reptiles. Special emphasis is given on the role of DNA barcoding in the identification and conservation of endangered reptile species. We also assessed the DNA barcoding scenario in all the four modern reptile orders—Crocodilia, Squamata, Testudines, and Sphenodontia. In conjugation with DNA barcoding, the amino acid informative sites within the barcode gene can be effectively used in the proper identification of different *Testudine* species. In contrast to other vertebrate groups, there is no DNA barcoding initiative exclusively for the reptiles, and there is a need to conduct more DNA barcoding studies on this important cold-blooded vertebrate group. The recent global

S. Trivedi (✉) · A. T. Aziz · C. Panneerselvam
Faculty of Science, Department of Biology, University of Tabuk, Tabuk, Kingdom of Saudi Arabia
e-mail: strivedi@ut.edu.sa

H. Rehman
Faculty of Science, Department of Biology, University of Tabuk, Tabuk, Kingdom of Saudi Arabia

Departments of Pathology, School of Medicine, The University of Alabama at Birmingham (UAB), Birmingham, AL, USA

S. Saggu
Faculty of Science, Department of Biology, University of Tabuk, Tabuk, Kingdom of Saudi Arabia

Departments of Dermatology, School of Medicine, The University of Alabama at Birmingham (UAB), Birmingham, AL, USA

S. K. Ghosh
Department of Biotechnology, Assam University, Silchar, Assam, India

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_18

barcoding initiative called “cold code” is expected to produce promising barcoding data on all cold-blooded vertebrates.

Keywords Biodiversity · Crocodilia · DNA barcoding · Reptiles · Squamata · *Sphenodontia* · *Testudines*

Abbreviations

ABBI	All birds barcoding initiative
BOLD	Barcode of life data system
CBOL	Consortium for the barcode of life
COI	Cytochrome c oxidase subunit 1
FISH-BOL	Fish barcode of life

1 Introduction

Reptiles are cold-blooded vertebrates which have horny scales and dry skins. The reptiles had dominated the world during the Mesozoic era which was regarded as the golden era of reptiles. During this era the giant dinosaurs prevailed which became extinct in the late Cretaceous period. Modern reptiles are categorized into four orders, namely, Crocodilia, Squamata, Testudines, and Sphenodontia. Poaching, overexploitation, habitat loss, risk from exotic species, emerging diseases, and effect of climate change are adversely affecting the reptile populations around the world (Oh 2006; Gibbons et al. 2000; Nijman et al. 2012). In some cases, application of rodenticide is also causing their decline by causing secondary poisoning (Kim et al. 2011). Besides these, road mortality is also one of the reasons for the decrease of vertebrates including the reptiles in our environment (Trombulak and Frissel 2000). The reptiles play a significant role both as predator and prey in the ecosystem. In agriculture, reptiles are necessary for biological control of rodents and insect pests.

Taxonomic classification and identification of reptiles is a challenging task. Generally, the reptiles are classified based on classical morphological approaches. In the last decade, the molecular concept of DNA barcoding has become a trend for identification and subsequent phylogenetic studies for different types of organisms ranging from the microbes to the mammals. The reptiles are no exception. In DNA barcoding, short standardized region of a gene is used as a molecular marker for species identification. About 650bp region from the 5' end of the *cytochrome c oxidase* subunit 1 (COI) gene was suggested as the barcode for animals (Hebert et al. 2003). Figure 1 shows the location of COI gene in the mitochondrial genome. This gene has been successfully used for barcoding different vertebrate groups like fishes (Trivedi et al. 2014; Ward et al. 2005), amphibians (Vences et al. 2012), reptiles (Nagy et al. 2012), birds (Aliabadian et al. 2013), and mammals (Alfonsi et al. 2013) and also for different invertebrate groups, like Porifera (Schuster et al. 2015), Cnidaria (Moura et al. 2007), Platyhelminthes (Larsson et al. 2008), Nemathelminthes (Derycke et al. 2008), Mollusca (Trivedi et al. 2012, 2013, 2015a), Arthropoda (Blanco-Bercial

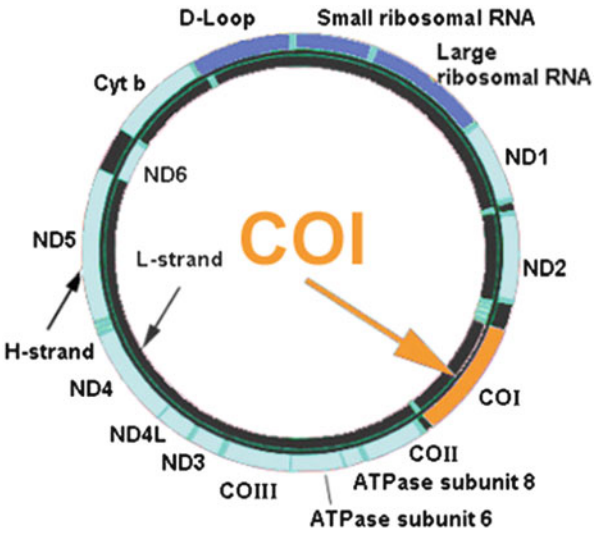


Fig. 1 Position of COI gene in the mitochondrial genome

et al. 2014; Trivedi et al. 2011; Bucklin et al. 2010), and Echinodermata (Uthicke et al. 2010; Ward et al. 2008). DNA barcoding has significant role in assessing marine biodiversity (Bucklin et al. 2011; Trivedi et al. 2015b).

In order to coordinate DNA barcoding globally, The Consortium for the Barcode of Life (CBOL) and Barcode of Life Data System—(http://www.barcodinglife.org)—were launched. International Barcode of Life Project (iBOL) has four central nodes, nine regional nodes, and seven country nodes spanning across the globe. A summary of the DNA barcoding procedure is shown in Fig. 2. In this review, we present the DNA barcoding scenario in different groups of reptiles.

2 DNA Barcoding Crocodilia

The order Crocodilia includes crocodiles, alligators, caimans, and gharials. Recently, a study based on both morphological and molecular methods was used to study the systematics of African crocodiles (Shirley et al. 2013). A separate barcoding study indicated that African crocodiles show higher genetic divergence than previously supposed (Schmitz et al. 2003). DNA Barcoding of Indian crocodiles showed the efficiency of this technique for species delimitation (Meganathan et al. 2013). For forensic purposes, skin samples were used for molecular identification of crocodiles (Eaton et al. 2010). There were taxonomic controversies regarding the phylogenetic position of African slender-snouted crocodile, *Crocodylus cataphractus*. Taxonomists had placed this species within the genus *Crocodylus* based on morphological characters. But recent molecular analysis revealed that it is not a member of the genus *Crocodylus* or *Osteolaemus* (McAliley et al. 2006). A time-calibrated species

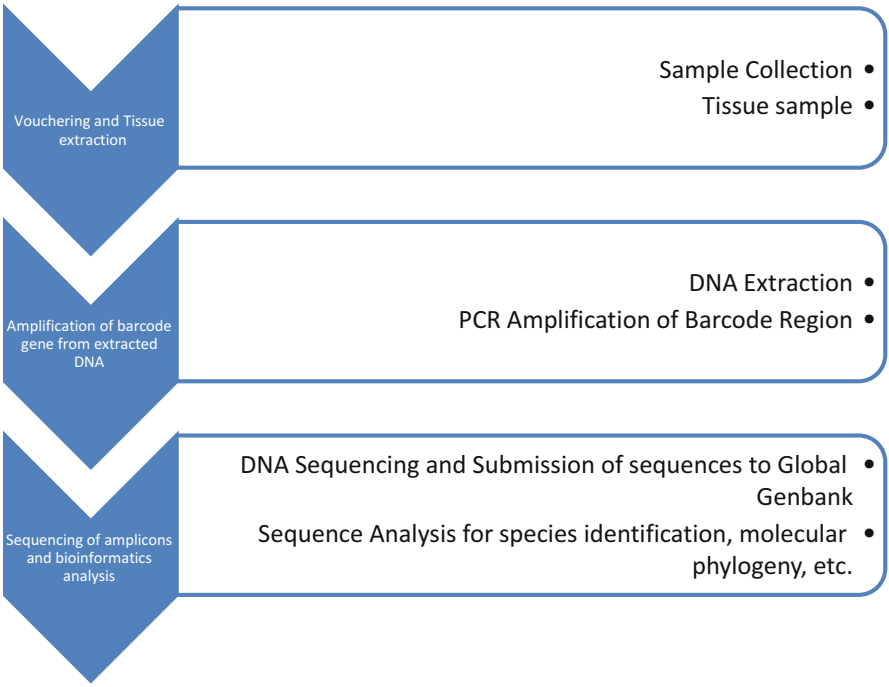

Vouchering and Tissue extraction
Sample Collection
Tissue sample
Amplification of barcode gene from extracted DNA
DNA Extraction
PCR Amplification of Barcode Region
Sequencing of amplicons and bioinformatics analysis
DNA Sequencing and Submission of sequences to Global Genbank
Sequence Analysis for species identification, molecular phylogeny, etc.

Fig. 2 Summary of DNA barcoding procedure

tree was constructed for *Crocodylus* which also revealed that there is higher diversity among *Crocodylus*. Another important finding of this study was that *Crocodylus* originated in Late Miocene Indo-Pacific from a common ancestor and spread to other parts of the globe (Oaks 2011).

3 DNA Barcoding Squamata

The order Squamata includes snakes, lizards, and worm lizards. In 1994 (long before the DNA barcoding concept was put forward in 2004), mitochondrial DNA was used to study phylogenetic relationship in a lizard *Sceloporus grammicus* (Arevalo et al. 1994). A large-scale DNA barcoding of reptiles including the Squamata was done using samples collected from Madagascar—a biodiversity hotspot (Nagy et al. 2012). This study revealed several new species which is one of objectives of DNA barcoding and also highlighted the role of DNA barcoding in conservation. Barcoding studies of squamate reptiles at the Comoros archipelago proved the efficiency of DNA barcoding in detecting even the isolated populations (Hawlitschek et al. 2013). DNA barcoding of German hepatofauna showed a success rate of almost 100% species identification (Hawlitschek et al. 2015). Confiscated snakes in Korea were identified by DNA barcoding (Jeong et al. 2013). Moreover, this technique is used to determine the venom samples (Pook and McEwing 2005) (Table 1).

Table 1 Publications on DNA barcoding of reptiles

Group	Title of article	References
Crocodilia	Rigorous approaches to species delimitation have significant implications for African crocodilian systematics and conservation	Shirley et al. (2013)
	Are crocodiles really monophyletic?—Evidence for subdivisions from sequence and morphological data	McAliley et al. (2006)
	Barcoding bushmeat: Molecular identification of Central African and South American harvested vertebrates	Eaton (2010)
	Identification of Indian crocodile species through DNA barcodes	Meganathan et al. (2013)
	A time-calibrated species tree of *Crocodylia* reveals a recent radiation of the true crocodiles	Oaks (2011)
Squamata	Comprehensive DNA barcoding of the herpetofauna of Germany	Hawlitschek et al. (2015)
	DNA barcode reference data for the Korean herpetofauna and their applications	Jeong et al. (2013)
	DNA barcoding assessment of genetic variation in two widespread skinks from Madagascar, *Trachylepis elegans* and *T. gravenhorstii* (Squamata: Scincidae)	Vences et al. (2014)
	DNA barcoding amphibians and reptiles	Vences et al. (2012)
	First large-scale DNA barcoding assessment of reptiles in the biodiversity hotspot of Madagascar, based on newly designed COI primers	Nagy et al. (2012)
	DNA-based identification of a snake in a wine bottle using universal primers: a case of mistaken identity	Gaur et al. (2012)
	Forensic wildlife parts and their product identification and individualization using DNA barcoding	Panday et al. (2014)
	Reliable DNA barcoding performance proved for species and island populations of Comoran Squamate reptiles	Hawlitschek et al. (2013)
	Mitochondrial DNA sequences from dried snake venom: a DNA barcoding approach to the identification of venom samples	Pook and McEwing (2005)
	Application of cytrochrome b DNA sequences for the authentication of endangered snake species	Wong et al. (2004)
	Mitochondrial DNA sequence divergence and phylogenetic relationships among eight chromosome races of the *Sceloporus grammicus* complex (Phrynosomatidae) in central Mexico	Arevalo et al. (1994)
Testudine	Taxonomic rank of Indian tortoise: revisit with DNA barcoding perspective	Kundu et al. (2013a)
	Amino acid analysis of cytochrome c oxidase subunit 1(COI) of Indian Testudines	Kundu et al. (2013b)
	Identification of commercialized turtle samples through DNA barcoding	Kundu et al. (2013c)
	DNA barcodes for globally threatened marine turtles: a registry approach to documenting biodiversity	Naro-Maciel et al. (2009)

(continued)

Table 1 (continued)

Group	Title of article	References
	The dazed and confused identity of Agassiz's land tortoise, *Gopherus agassizii*, the description of a new species, and its consequences for conservation	Murphy et al. (2011)
	DNA barcoding of Brazilian sea turtles (Testudines)	Vargas et al. (2009)
	Phylogenetic relationship among extinct and extant turtles: the position of Pleurodira and the effects of the fossils on rooting crown-group turtles	Sterli (2010)
	Comparing and combining distance-based and character-based approaches for barcoding turtles	Reid et al. (2011)
	Multiple data sets, high homoplasy, and the phylogeny of softshell turtles (Testudines: Trionychidae)	Engstrom et al. (2004)
	Mitochondrial DNA sequences suggest a revised taxonomy of Asian flapshell turtles (*Lissemys* Smith, 1931) and the validity of previously unrecognized taxa (Testudines: Trionychidae)	Praschag et al. (2011)
Sphenodontia	A new rhynchocephalian from the late Jurassic of Germany with a dentition that is unique among Tetrapods	Rauhut et al. (2012)

A fascinating case relating to wildlife forensic science was revealed with the confiscation of a snake—wine bottle from a bar in Bangalore, India. Initially, based on morphological characters, the snake was identified as Indian cobra—*Naja naja*. But the subsequent barcoding analysis revealed that the snake was Indo-Chinese spitting cobra—*Naja siamensis* (Gaur et al. 2012). The utility of DNA barcoding in wildlife forensics is noteworthy (Panday et al. 2014; Vargas et al. 2009). Cytochrome b DNA sequences were used for DNA barcoding of endangered snakes (Wong et al. 2004). The location of Cytochrome b (Cyt b) which is adjacent to the D-loop (displacement loop) in the mitochondrial genome is shown in Fig. 1.

4 DNA Barcoding Testudines

Testudines includes turtles and tortoises. They have shelled reptiles. One of the main differences between turtles and tortoises is that turtles are aquatic but tortoises are terrestrial. Recent barcoding studies have shown that Agassiz's desert tortoise *Gopherus agassizii* occupies only 30% of its initially proposed range, and in the past 30 years, there is a drastic decline in their population. If proper conservation strategies are not implemented, this endangered species may face extinction (Murphy et al. 2011).

A study on the sea turtles of Brazil showed that COI sequences could be efficiently used to barcode these species (Vargas et al. 2009). DNA barcoding is very effective for identification, evaluation, and conservation of threatened marine turtles (Naro-Maciel et al. 2009). The efficacy of this technique was also tested with

several tortoise species of Northeast India (Kundu et al. 2013a). A nuclear intron from the RNA fingerprint protein 35 was first time used to study turtle phylogeny (Fujita et al. 2004).

Rhodin et al. (2010) provided an updated checklist of the world turtles. Softshell turtles (Trionychids) are facing an environmental and anthropogenic threat which is leading to the decline in its population (Gong et al. 2006). Due to complex morphological data in the different life cycle stages, it becomes challenging to identify the softshell turtles. This is a serious hurdle for the taxonomists and conservation strategists. DNA barcoding provided clarity in the species discrimination of softshell turtles (Kundu et al. 2013b). The ambiguities regarding the species identification of turtles and tortoises can be solved by DNA barcoding (Reid et al. 2011). Cytochrome c oxidase subunit 1 (COI) is an ideal mitochondrial marker for phylogenetic studies as it is most abundant among the three cytochrome oxidase subunits and amino acid sequences have highly conserved functional domains and variable sites (Morlais and Severson 2002).

After that it has been revealed that along with the COI sequences, the amino acids informative sites within the COI region may play a vital role in discriminating the different *Testudine* species (Kundu et al. 2013c). This study on Indian *Testudines* showed that first codon position had GC bias (54%) whereas the second and third codon positions had AT bias of 56.4% and 67.4%, respectively. Furthermore, it was found that nucleotide variation and amino acid variation were found in 293 and 57 positions, respectively.

5 DNA Barcoding Sphenodontia

Tautaras of New Zealand are included in the order Sphenodontia. *Sphenodon* is the only surviving genus belonging to the order Sphenodontia. It is speculated that they lived along with the dinosaurs and separated from other reptile groups more than 200 million years ago. A team of scientists regards them as living fossils. About a quarter of *Sphenodon* population have become extinct in the last century (Daugherty et al. 1990).

Scientists have reported that bad taxonomy can lead to the severe decline of species (Daugherty et al. 1990; May 1990). Jones et al. (2008) studied the pre-Pleistocene and post-Mesozoic rhynchocephalian fossil remains from the Miocene of New Zealand. This remains dating back 18 million years, has raised questions whether New Zealand was fully submerged some 25 million years ago as previously thought. A new fossil relative of Sphenodon from the latest Jurassic of southern Germany, *Oenosaurus muehlheimensis*, has been described (Rauhut et al. 2012). Scientists at Griffith University are working on the mitochondrial genome of Sphenodon.

6 Conclusion

Although reptiles represent significant vertebrate diversity, fewer barcoding studies have been conducted compared to other vertebrates. At present, there is no global initiative for barcoding exclusively on reptiles. In contrast, we have global barcoding initiatives for other vertebrates like fish (FISH-BOL—http://www.fishbol.org), birds (All Birds Barcoding Initiative—ABBI), and mammals (Mammalia Barcode of Life—http://www.mammaliabol.org). Our experience in the lab has shown that different sets of primers or cocktail of primers are needed for barcoding reptiles which makes barcoding of reptiles relatively difficult. This technical problem has been highlighted, and subsequently, new primer sets were suggested (Nagy et al. 2012).

Recently, a global effort to barcode all cold-blooded vertebrates has been launched which is called "cold code." This initiative is expected to reveal essential barcoding data on cold-blooded vertebrates including reptiles (Murphy et al. 2011). Paleobarcoding is supposed to decipher essential insights into the identification and phylogeny of ancient reptiles. With the advent of high-throughput DNA sequencing, it is now possible to amplify mass collection of different species and even environmental DNA. Metabarcoding can shorten the timespan for identification procedure. With these advancements, DNA barcoding can serve as an important molecular tool for biodiversity assessment and conservation of reptiles.

Acknowledgments We acknowledge the Deanship of Scientific Research, University of Tabuk, Saudi Arabia, for the support provided through the Project No. S-1436-0252. We also express our gratitude to the Saudi Digital Library (SDL).

Conflict of Interest No conflict of interest reported in the current work.

References

Alfonsi E, Me ME, Fuchs S (2013) The use of DNA barcoding to monitor the marine mammal biodiversity along the French Atlantic coast. In: Nagy ZT, Backeljau T, De Meyer M, Jordaens K (eds) DNA barcoding: a practical tool for fundamental and applied biodiversity research, ZooKeys, vol 365, pp 5–24

Aliabadian M, Beentjes KK, Roselaar CS (2013) DNA barcoding of Dutch birds. In: Nagy ZT, Backeljau T, De Meyer M, Jordaens K (eds) DNA barcoding: a practical tool for fundamental and applied biodiversity research, ZooKeys, vol 365, pp 25–48

Arevalo E, Davis SK, Sites JW (1994) Mitochondrial DNA sequence divergence and phylogenetic relationships among eight chromosome races of the *Sceloporus grammicus* complex (Phrynosomatidae) in central Mexico. Syst Biol 43:387–418

Blanco-Bercial L, Cornils A, Copley N, Bucklin A (2014) DNA barcoding of marine Copepods: assessment of analytical approaches to species identification. PLoS Curr 6. https://doi.org/10.1371/currents.tol.cdf8b74881f87e3b01d56b43791626d2

Bucklin A, Ortman BD, Jennings RM, Nigro LM, Sweetman CJ et al (2010) A "Rosetta Stone" for metazoan zooplankton: DNA barcode analysis of species diversity of the Sargasso Sea (Northwest Atlantic Ocean). Deep-Sea Res II 57:2234–2247

Bucklin A, Steinke D, Blanco-Bercial L (2011) DNA barcoding of marine Metazoa. Annu Rev Mar Sci 3: 471 47171 3: 471tazoa. https://doi.org/10.1146/annurev-marine-120308-080950

Daugherty CH, Cree A, Hay JM, Thompson MB (1990) Neglected taxonomy and continuing extinctions of tuatara (Sphenodon). Nature 347:177–179

Derycke S, Fonseca G, Vierstraete A, Vanfleteren J, Vincx M, Moens T (2008) Disentangling taxonomy within the Rhabditis (Pellioditis) marina (Nematoda, Rhabditidae) species complex using molecular and morphological tools. Zool J Linnean Soc 152:1–15

Eaton MJ, Meyers GL, Kolokotronis SO, Leslie MS, Martin AP, Amato G (2010) Barcoding bushmeat: molecular identification of Central African and South American harvested vertebrates. Conserv Genet 11:1389–1404

Engstrom TN, Shaffer HB, McCord WP (2004) Multiple data sets, high homoplasy, and the phylogeny of softshell turtles (Testudines: Trionychidae). Syst Biol 53(5):693–710

Fujita MK, Engstrom TN, Starkey DE, Shaffer HB (2004) Turtle phylogeny: insights from a novel nuclear intron. Mol Phylogenet Evol 31:1031–1040

Gaur A, Singh CS, Sreenivas A, Singh L (2012) DNA-based identification of a snake in a wine bottle using universal primers: a case of mistaken identity. Forensic Sci Int 214:e51–e53

Gibbons JW, Scott DE, Ryan TJ et al (2000) The global decline of reptiles, deJa vu amphibians. Bioscience 50:653–666

Gong S, Wang J, Shi H, Song R, Xu R (2006) Illegical trade and conservation requirements of freshwater turtles in Nanmao, Hainan province, China. Oryx 40:331–336

Hawlitschek O, Nagy ZT, Berger J, Glaw F (2013) Reliable DNA barcoding performance proved for species and island populations of comoran squamate reptiles. PLoS One 8(9):e73368. https://doi.org/10.1371/journal.pone.0073368

Hawlitschek O et al (2015) Comprehensive DNA barcoding of the herpetofauna of Germany. Mol Ecol Resour. https://doi.org/10.1111/1755-0998.12416

Hebert PD, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proc Biol Sci 270(1512):313–321

Jeong TJ et al (2013) DNA barcode reference data for the Korean herpetofauna and their applications. Mol Ecol Resour. https://doi.org/10.1111/1755-0998.12055

Jones MEH, Worthy TH, Tennyson AJD, Evans SE (2008) The first pre-Pleistocene record of a tuatara (Sphenodon)-like animal from New Zealand and implications for the Oligocene drowning. J Vertebr Paleontol 28:97–98

Kim JB, Min MS, Park DS, Song JY (2011) Red Data Book of Endangered Amphibians and Reptiles in Korea. National Institute of Biological Resources, Incheon, South Korea

Kundu S, Das KC, Ghosh SK (2013a) Taxonomic rank of Indian tortoise: revisit with DNA barcoding perspective. DNA Barcodes 1:39–45. https://doi.org/10.2478/dna-2013-0003

Kundu S, Das KC, Ghosh SK (2013b) Amino acid analysis of cytochrome c oxidase subunit 1(COI) of Indian testudines. J Environ Sociobiol 10(1):43–48

Kundu S, Das KC, Ghosh SK (2013c) Identification of commercialized turtle samples through DNA barcoding. J Environ Sociobiol 10(1):37–42

Larsson K, Ahmadzadeh A, Jondelius U (2008) DNA taxonomy of Swedish Catenulida (Platyhelminthes) and a phylogenetic framework for catenulid classification. Org Divers Evol 8:399–412

May RM (1990) Taxonomy as destiny. Nature 347:129–130

McAliley LR, Willis RE, Ray DA, White PS et al (2006) Are crocodiles really monophyletic? Evidence for subdivisions from sequence and morphological data. Mol Phylogenet Evol 39 (1):16–32

Meganathan PR, Dubey B, Jogayya KN, Haque I (2013) Identification of Indian crocodile species through DNA barcodes. J Forensic Sci 58(4):993–998

Morlais I, Severson DW (2002) Complete mitochondrial DNA sequence and amino acid analysis of the cytochrome c oxidase subunit 1 (COI) from *Aides aegypti*. DNA Seq 13:123–127

Moura CJ, Harris DJ, Cunha MR, Rogers AD (2007) DNA barcoding reveals cryptic diversity in marine hydroids (Cnidaria, Hydrozoa) from coastal and deep-sea environments. Zool Scr 37 (1):93–108

Murphy RW, Berry KH, Edwards T et al (2011) The dazed and confused identity of Agassizsiz. The dazise, *Gopherus agassizii*, the description of a new species, and its consequences for conservation. ZooKeys 113:39–71

Nagy ZT, Sonet G, Glaw F (2012) First large-scale DNA barcoding assessment of reptiles in the biodiversity hotspot of Madagascar, based on newly designed COI primers. PLoS One 7(3): e34506

Naro-Maciel E, Reid B, Fitzsimmons NN (2009) DNA barcodes for globally threatened marine turtles: a registry approach to documenting biodiversity. Mol Ecol Resour 10:252–263

Nijman V, Shepherd CR, Mumpuni SKL (2012) Over-exploitation and illegal trade of reptile in Indonesia. Herpetol J 22:83–89

Oaks JR (2011) A time-calibrated species tree of Crocodylia reveals a recent radiation of the true crocodiles. Evolution. https://doi.org/10.1111/j.1558-5646.2011.01373.x

Oh JG (2006) The herpatofauna of Jeju island. In: Chung SH, Oh JG, Kim WB (eds) The fauna of Mr. Halla. Halla Institute of Ecology and Culture, Jeju, South Korea, pp 23–71

Panday R, Jha DK, Thapa N, Pokharel BR, Aryal NK (2014) Forensic wildlife parts and their product identification and individualization using DNA barcoding. Open Forensic Sci J 7:6–13

Pook CE, McEwing R (2005) Mitochondrial DNA sequences from dried snake venom: a DNA barcoding approach to the identification of venom samples. Toxicon 46:711–715

Praschag P, Stuckas H, Packert M, Maran J, Fritz U (2011) Mitochondrial DNA sequences suggest a revised taxonomy of Asian flapshell turtles (Lissemys SMITH, 1931) and the validity of previously unrecognized taxa (Testudines: Trionychidae). Vertebr Zool 6(1):147–160

Rauhut OWM, Heyng AM, López-Arbarello A, Hecker A (2012) A new rhynchocephalian from the Late Jurassic of Germany with a dentition that is unique amongst tetrapods. PLoS One 7(10): e46839. https://doi.org/10.1371/journal.pone.0046839

Reid BN, Le M, McCord WP, Iverson JB, Georges A, Bergmann T, Amato G, Desalle R, Naro-Maciel E (2011) Comparing and combining distance-based and character-based approaches for barcoding turtles. Mol Ecol Resour 11:956–967

Rhodin AJ, van Dijk PP, Iverson JB, Shaffer HB (2010) Turtles of the world, 2010 update: annotated checklist of taxonomy, synonymy, distribution and conservation status. Chelonian Res Monogr 5:000.85–000.1 64

Schmitz A, Mansfeld P, Hekkala E, Shine T et al (2003) Molecular evidence for species level divergence in African Nile Crocodiles *Crocodylus niloticus* (Laurenti, 1786). C R Palevol 2:703–712

Schuster A, Erpenbeck D, Pisera A, Hooper J, Bryce M, Fromont J et al (2015) Deceptive desmas: molecular phylogenetics suggests a new classification and uncovers convergent evolution of lithistid demosponges. PLoS One 10(1):e116038. https://doi.org/10.1371/journal.pone.0116038

Shirley MH, Vliet KA, Carr AN, Austin JD (2013) Rigorous approaches to species delimitation have significant implications for African crocodilian systematics and conservation. Proc R Soc B 281:20132483

Sterli J (2010) Phylogenetic relationships among extinct and extant turtles: the position of Pleurodira and the effects of the fossils on rooting crown-group turtles. Contrib Zool 79 (3):93–106

Trivedi S, Ghosh SK, Choudhury A (2011) Cytochrome c oxidase subunit 1 (COI) sequence of *Macrobrachium rosenbergii* collected from Sunderbans. J Environ Sociobiol 8(2):169–172

Trivedi S, Ghosh SK, Choudhury A (2012) Mitochondrial DNA sequence of Cytochrome c oxidase subunit 1 (COI) region of an oyster *Crassostrea cuttakensis* collected from Sunderbans. J Environ Sociobiol 9(1):13–16

Trivedi S, Ghosh SK, Choudhury A (2013) DNA sequence of Cytochrome c oxidase subunit 1 (COI) region of an oyster *Saccostrea cucullata* collected from Sunderbans. J Environ Sociobiol 10(1):77–81

Trivedi S, Affan R, Alessa AHA et al (2014) DNA barcoding of Red Sea fishes from Saudi Arabia – the first approach. DNA Barcodes 2:17–20

Trivedi S, Aloufi AA, Ansari AA, Ghosh SK (2015a) Molecular phylogeny of oysters belonging to the genus Crassostrea through DNA barcoding. J Entomol Zool Stud 3(1):21–26

Trivedi S, Aloufi AA, Ansari AA, Ghosh SK (2015b) Role of DNA barcoding in marine biodiversity assessment and conservation: an update. Saudi J Biol Sci 23:161–171. https://doi.org/10.1016/j.sjbs.2015.01.001

Trombulak CS, Frissel AC (2000) Review of ecological effects of roads on terrestrial and aquatic communities. Conserv Biol 14(18):14. https://doi.org/10.1046/j.1523-1739.2000.99084.x

Uthicke S, Byrne M, Conand C (2010) Genetic barcoding of commercial Bêche-de-mer species (Echinodermata: Holothuroidea). Mol Ecol Resour 10(4):634–646. https://doi.org/10.1111/j.1755-0998.2009.02826.x

Vargas SM, Arauas FCF, Santos FR (2009) DNA barcoding of Brazilian sea turtles (Testudines). Genet Mol Biol 32(3):608–612

Vences M, Nagy ZT, Sonet G, Verheyen E (2012) DNA barcoding amphibians and reptiles. In: Kress WJ, Erickson DL (eds) DNA barcodes: methods and protocol. Springer

Vences M, Lima A, Miralles A, Glaw F (2014) DNA barcoding assessment of genetic variation in two widespread skinks from Madagascar, *Trachylepis elegans* and *T. gravenhorstii* (Squamata: Scincidae). Zootaxa 3755(5):477–484

Ward RD, Zemlak TS, Innes BH (2005) Barcoding Australia's fish species. Philos Trans R Soc Lond B 360:1847–1857

Ward RD, Holmes BH, O'hara TD (2008) DNA barcoding discriminates echinoderm species. Mol Ecol Resour 8(6):1202–1211

Wong KL, Wang J, But PP, Shaw PC (2004) Application of cytrochrome b DNA sequences for the authentication of endangered snake species. Forensic Sci Int 139:49–55

DNA Barcoding in Avian Species with Special Reference to Taxonomically Wide Biogeographic Studies

Farhina Pasha

Abstract The establishment of DNA Barcode Project in 2003, intending to construct a strong molecular identification tool via standardized genetic sequences, marked a new era of species identification and taxonomy. DNA barcoding so far proved to be simple and one of the excellent tools for identification of not only animals and plants but also the aves. The avian fauna represents an excellent contender for testing DNA barcode validation as aves or birds are amongst the most prominent groups in which a wide variety of morphological, genetic and behavioural studies have been conducted there by establishing a prime line of stable taxonomy . The idea of All Bird Barcoding Initiative (ABBI) was conceived in 2005 with the intention to collect genetic data samples for deciphering DNA barcode for over 10,000 known avian species. Regardless of hundreds of vigilant studies carried out during the past decade, there are still numerous avian species to be discovered and identified. ABBI is a new hope for a speedy identification of novel avian species and will also help in hundreds of new samples to be identified thereby opening new avenues for avian identification and its related scientific research. Adding on, with the advancement of mt-DNA gene cytochrome *c* oxidase I (COI) library via DNA barcoding projects of avian species, there will be better understanding of different avian realms and taxonomic territories. It will also serve as an unbiased taxonomic representation of different avian groups. The advantage is that the DNA barcode sequences deposited in these databases are of high quality and are standardized, therefore has fewer ambiguities, short sequence span, bidirectional sequencing and uniform sequence alignment. It has been observed that the rate of error on BOLD is much lower than on other databases DNA barcoding in avian species in future will undoubtedly provide more specific species identification, recognition of cryptic species and tracing the line of avian evolution through different eras, deciphering their causes of divergence.

F. Pasha (✉)
Faculty of Sciences, Department of Biology, University of Tabuk, Tabuk, Saudi Arabia
e-mail: fbasha@ut.edu.sa

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_19

Keywords Avian species · DNA barcoding · COI gene library · Island birds · Sea birds · Migratory birds and continental birds

1 Introduction

The avian fauna represents an excellent contender for testing DNA barcode validation. Aves or birds are amongst the most prominent groups in which a wide variety of morphological, genetic and behavioural studies have been conducted to establish a prime line of stable taxonomy (Baker et al. 2009; Gill 2007). Carl Woese was the first researcher who applied the nucleotide sequence variation in a sole gene to trace the evolutionary relationships. On the basis of his research, he inferred that sequences differ in a conserved gene and the ribosomal RNA can be used to trace phylogenetic interactions and sequences as that revealed in *Archaea* furthermore leading to trace tree of life (Woese and Fox 1977). As a recent advance, the polymerase chain reaction has emerged as an effective tool to examine sequence diversity in any gene. It is evident that the genes like rRNA-encoding genes evolve slowly. These genes are closely related and are crucial in tracing primaeval relationships, whereas the genes that evolve rapidly may overwrite, but they do divulge divergence amongst closely related species (Woese 2000).

The sequence divergence in mitochondrial DNA (mtDNA) can be utilized to outline evolutionary history within the species was first revealed by John Avise. This can potentially be used for linking systematic and population genetics thereby establishing phylogeography (Avise et al. 1987). Though some species show phylogeographic subdivisions, but they are frequently in close proximity "at distances much shorter than the internodal branch lengths of the species tree" as reported by Moore (1995). Therefore it can be concluded that the sequence divergences within the species is larger and, mtDNA confines the discontinuities in gene identified by taxonomists as species and because of this reason, taxonomic revisions use analysis of mtDNA divergences regularly. This has played a major part in identification of native avian species on the basis of mtDNA divergences (Avise and Walker 1998; Gill and Slikas 1992; Murray et al. 1994; AOU 1998; Banks et al. 2000, 2002, 2003).

2 All Birds Barcoding Initiative (ABBI)

The idea of All Bird Barcoding Initiative (ABBI) was conceived in 2005 with the intention to collect genetic data samples for deciphering DNA barcode for over 10,000 known avian species. Regardless of hundreds of vigilant studies carried out during the past decade, there are still numerous avian species to be discovered and identified. ABBI is a new hope for a speedy identification of novel avian species and will also help in hundreds of new samples to be identified thereby opening new avenues for avian identification and its related scientific research.

The establishment of DNA Barcode Project in 2003, intending to construct a strong molecular identification tool via standardized genetic fragment, marked a new era of species identification and taxonomy. The library of sequences so formed has reached almost five million sequences in 2015 (formally projected in 2009; Frézal and Leblois 2008). There are more than 47,000 sequences of 6000 avian species deposited here on the Barcode of Life Data (BOLD) (available on www.boldsystems.org; access 2016) from the entire worldwide avian diversity of 10,473 species (Clements et al. 2015). ABBI stands as the first global campaign for obtaining a specific taxonomic group that of avian species; 34,000 avian DNA barcode records are available here with sequences of over 4280 species and 37 orders (Baker et al. 2005; Stoeckle 2005).

Avian species were amongst the first group selected for the testing of DNA barcoding because of their well-established taxonomy which made it possible to recognize the genetic discrepancy and Linnaean specie restrictions (Hebert et al. 2004). The growing ABBI open- access electronic library tends to link DNA barcodes, different specimen and associated data collections by depositing records in BOLD/GenBank/EMBL/DDJB. This library will definitely be an important source for conservation management, biodiversity experts' ornithologists, ecologist's, public health officials and interested common public. Not only this, it will serve as a benchmark for identification of birds that strike the aeroplanes, helping airline safety and also for identification and tracking divergence in migratory avian species (other animal and plant life). In the future, these samples can be analysed regardless of their size, age, plumage or sex.

3 Modus Operandi

It has been well established now that 648 bp region at the 5′ end of the mitochondrial DNA gene cytochrome *c* oxidase subunit 1 (COI or cox1) is ideal for most of the animal species including avian species (Hebert et al. 2003, 2004). Primarily the samples available for DNA barcoding in avian species include fresh tissues of skin and skeleton, blood or feathers, most of which are conserved at low temperature or in ethanol. In case of historical samples, vouchered museum specimens are the source of sample which usually includes skin and skeleton for various biochemical and DNA analysis as a recent practice (Arctander and Fjeldsa 1994; Seutin et al. 1991). Though the historical samples which usually are over two decades old are not so well preserved for DNA analysis, highly degraded DNA is obtained (Lindahl 1993). These samples are very difficult to amplify and to be used for further studies (Zimmermann et al. 2008). The solution to this has occurred by selecting a 100–200 bp fragment of the initially projected ~650 bp COI region known as the mini-barcode (Hajibabaei et al. 2006). However the success rate is very low as it works merely in the restricted taxonomic group. An alternate option is the use of DNA repair enzyme which increases the success rate of amplification especially when the DNA is damaged by ageing or chemicals/preservatives (Evans 2007), the

most common being oxidative and hydrolytic damage (Hoss et al. 1996). By targeting a small overlapping region from any section of DNA, an entire sequence can be gained from an ancient historical vouchered sample, and hence DNA barcoding in avian species will be possible by a set of conserved primers which allow proper amplification of COI barcoding region (Millar et al. 2008).

To outline the pattern of COI sequence divergence in avian species, the initial sample includes single individual selected on the basis of convenience rather than taxonomic concern to determine the COI divergence between species. At the next level, one to three or more individuals are examined to provide information about intraspecific sequence divergence. Thereby to obtain COI divergence at each level, the number of individuals is increased, and a sample size of 20 years can be analysed by this method. A 2% sequence divergence was observed by Paul amongst individuals in *Birds of North America* using the above methodology.

Major difficulty encountered in identification of species by mitochondrial DNA studies is the interference of nuclear mitochondrial DNA segments (NUMT) or pseudogenes, as they are non-coding sequences which are accumulated via random mutations (Kerr 2010; Stoeckle et al. 2012). Although the NUMT have a stop codon and idles which makes it easy for their identification and deletion from the database, there are some cryptic pseudogenes that lack this characteristics, and a slight genetic divergence is seen because of them instead of the original COI sequence (Stoeckle et al. 2012). In avian species they are far more problematic as their common primers show least resemblance to mitochondrial gene. *Tyrannidae* are a classic example, these are large family of Passeriforme order, and North, Central and South Americas' birds are the prime representatives of this order. Cryptic pseudogenes also occur as a result of amplification done via blood sample having high number of nuclear DNA copies related to mitochondrial segments (Kerr 2010). As NUMT are placed in nuclear DNA, they tend to have low evolutionary rate than mitochondrial COI (mtCOI) gene, which is 2–14 times greater as reported by Kerr (2010), where NUMT sequences were found to be 14.2%. Even though the COI pseudogenes show their presence occasionally, it may delude to the assumption of cryptic divergence and split conspecific in an additional Barcode Index Number (BIN). This problem can be solved by supplementing with the other barcodes especially the nuclear DNA barcode (Dasmahapatra and Mallet 2006).

4 DNA Barcoding in Island Birds

Avian species native to various islands have always attracted ornithologist all over the world. They attract the attention of researchers, interested in evolution and biogeographic studies globally. These aves are exceptionally unique in their behavioural patterns and are known for their reproductively isolated behaviour from the continental avian population. They also have a very specific colony pattern (Newton 2003). Though most of the avian species have nuclear ancestral origin, many groups show a spectacular adaptive variation in food habitats, colony patterns and of some

extent structural patterns too. As a consequence a diverse group of closely related avian species have occurred, occupying wider habitats. There are many studies conducted on island avian species, and their population studies have been undertaken to understand their biogeographic orientation via DNA barcoding (Lohman et al. 2010; Campagna et al. 2012; Nishiumi and Kim 2015). A phylogenetic tree for identifying the cases of reverse colonization amongst island aves was constructed by Nishiumi and Kim (2015) by obtained DNA sequences of Holarctic avian species (i.e. from Japanese islands to mainland Asia) from BOLD and by perceiving the topology in relation to geographic origin of the samples. The study identified 5 species with strong reverse colonization pattern amongst the 118 samples, whereas 39 avian species had their genetic makeup as that of continental origin but were found breeding on Japanese islands. Overall there is still a great need of studies to be undertaken for deciphering and obtaining more genetic markers especially of the remote island avian species, a wider sampling population and more ordain phylogenetic reconstruction for, e.g. maximum likelihood and Bayesian analysis methods. The estimation of divergence time between the continental and island avian species also needs to be studied for obtaining accuracy in reverse colonization process.

5 DNA Barcoding in Continental Species

As compared to the island birds, the continental species exhibit more closely related genetic patterns exhibiting a familiar geological history amongst them (Newton 2003). Several studies had been undertaken to establish the various diversifying factors using COI variations framework (Johnsen et al. 2010; Lijtmaer et al. 2011; Tavares et al. 2011; Milá et al. 2012; Chaves et al. 2015).

DNA barcoding can reliably identify the closely related sister species of birds by using COI sequence. To test the efficacy of DNA barcodes for the identification of closely related species, Tavares and Baker (2008) compared *COI* in 60 (~650 bp) sister-species pairs from 10 orders of birds. Individuals of each species were of monophyly in a neighbour-joining (NJ) tree in all the pairs; each species was possessing fixed mutational differences differentiating them from their sister species (Tavares and Baker 2008).

In a study Tavares et al. (2011) joined the DNA barcode of Brazilian avian species with those of Argentina to study the genetic structure of Neotropical species. They observed 75% species had low intraspecific COI divergence, and the number of avian species in this territory was having outsized genetic splits than in North American birds. It is more likely, that the effect of glacial cycles, isolating populations in North America has generated speciation (Lijtmaer et al. 2011). In another study undertaken by Neotropics (Milá et al. 2012), genetic diversification amongst eastern and western Amazon population was observed, and large level of intraspecific divergence was reported. This genetic deviation amongst the population was not equivalent to phenotypic variation in plumage colouration (Milá et al. 2012). Another study conducted by Chaves et al. (2015) using Bayesian tree reconstructions of DNA

barcode sequence for avian species of the Brazilian Cerrado and the Atlantic Forest showed 10.4% were non-monophyletic. It is well indicated by the fact that from 2004 to 2016, almost 116 articles have been published on PubMed (search criteria, DNA barcoding in avian species). This definitely has marked a new era in bird identification and their taxonomic coverage, but credentials of cryptic species and diversifying factors still need to be deciphered for most of the geographical realms.

6 DNA Barcoding in Migratory Birds

Most of the avian species have high metabolic rate thereby requiring plenteous food supply throughout. But this supply may not occur throughout the year, in many geographic realms; therefore birds have evolved and adapted very effective method of travelling interminability to food-rich grounds amid great energy-conserving efficiency. Though the characteristics of migratory birds do not differ totally from the non-migratory birds, many transitional groups exist amid them. These characteristics can be seen in a single local population exhibiting partial migration. The significance in studying these migratory birds lies in the fact that birds are very sensitive to environmental changes therefore serving as very good biological indicators. Majority of them are at peak of their decline because of loss of habitat, immense hunting, introduction of foreign species, climatic changes resulting in disturbed rainfall pattern and high temperature. DNA Barcoding has played a significant role in species identification, but in the case of migratory birds, it adds on to classify the distinct lineages and discovering new species. Not only this, but through DNA barcoding, many cryptic species have been identified which previously were classified as single species (Hajibabaei et al. 2005). DNA barcoding also helps in characterizing the inter- and intraspecific genetic diversity which is very important in governing the distinct lineage in migratory birds. Many studies have been conducted using DNA barcoding in migratory birds (Table 1). The researchers from many countries such as North America, Korea, Turkey, Argentina and Scandinavia have extensively worked using DNA barcoding as effective tool for species identification and to visualize the patterns of distinct lineages (Hebert et al. 2004; Yoo et al. 2006; Kerr et al. 2009; Johnsen et al. 2010).

The basis protocol followed for DNA barcoding in birds/migratory birds begins with the collection of blood sample taken via brachial vein. This is followed by extraction of DNA commonly using Genomic DNA Extraction Kits. Cytochrome oxidase I gene is usually used as molecular marker as it is accepted by Consortium for Barcode of Life Database (CBOL) and ABBI as a unique DNA barcode marker for avian species. Standard primers are used for PCR amplification Bird F1 (TTCTCCAACC ACAAAGACATTGGCAC), Bird R1 (AC GT GGGA GATAATTCCAAATCCTG) and Bird R2 (ACTACATGTGA GATG ATTCC GAA TCCAG) (Johnsen et al. 2010). Whereas protocol for COI amplification is in accordance with Hebert et al. (2004). The final PCR product is then subjected to commercial sequencing, and finally all obtained sequences are submitted to Consortium for Barcode of Life Database (BOLD). A

Table 1 Recent publications on DNA barcoding in migratory birds

S. No.	Title of publication	References	Search criteria
1	DNA Barcoding of Birds at a Migratory Hotspot in Eastern Turkey Highlights Continental Phylogeographic Relationships	Bilgin et al. (2016)	DNA barcoding in migratory birds
2	Overseas seed dispersal by migratory birds	Viana et al. (2016)	DNA barcoding in migratory birds
3	Avian haemosporidians from Neotropical highlands: Evidence from morphological and molecular data	González et al. (2015)	DNA barcoding in migratory birds
4	Low resolution of mitochondrial COI barcodes for identifying species of the genus Larus (Charadriiformes: Laridae)	Kwon et al. (2012)	DNA barcoding in migratory birds
5	A DNA microarray for identification of selected Korean birds based on mitochondrial cytochrome c oxidase I gene sequences	Chung et al. (2010)	DNA barcoding in migratory birds
6	DNA barcoding techniques for avian influenza virus surveillance in migratory bird habitats	Lee et al. (2010)	DNA barcoding in migratory birds
7	Application of DNA barcoding technique in avian influenza virus surveillance of wild bird habitats in Korea and Mongolia	Lee et al. (2010)	DNA barcoding in migratory birds

Search source: https://www.ncbi.nlm.nih.gov/pmc/?term=DNA+barcoding+in+migratory+birds

neighbour-joining tree is then prepared by combining sample and BOLD sequences, and intraspecific and interspecific distance is calculated (many software like MEGA 4.0, TSC 1.21, etc. are currently available for this).

7 DNA Barcoding in Sea Birds

Sea birds form unique class in itself. They represent avian species, which are prominent sea dwellers, but cannot be included in sea fauna. They spend most of their life in marine environment, obtaining food from water and for reproduction and colonization depending on the sandy beaches. Netherland is categorized as one of the best countries for ornithological studies where approximately 147 species of birds (including sea birds) have been DNA barcoded (Aliabadian et al. 2013). Birds are extremely sensitive to environmental changes especially the sea birds because global warming is causing the continuous rise of sea level and sea salinity, which can play a crucial role in the extinction of many species of sea birds, so DNA barcoding can help to keep the record of endangered species. So there is an urgent need for more studies in sea birds using DNA barcoding for their conservation.

8 DNA Barcoding for the Identification of Illegal Trades of Birds

DNA barcoding is not only used for the identification of species of particular geographic area but also used to identify the illegal trades of birds.

Illegal wildlife trade imposes a negative impact on the survival of a species as it introduces the invasion of other species and the pathogens (Rosen and Smith 2010). Molecular markers play important role in the identification of the forensic samples. Parrots and cockatoos (*Psittaciformes*) are charming birds; their plumage makes them highly enviable pets. The illegal trade in parrots and cockatoos is a serious issue and poses a threat to the viability of local populations, and their transport to nonendemic areas may genetically pollute the avifauna (Coghlan et al. 2012).

Priscila F. M and co-workers identified a case were a man was caught in a Brazilian airport who was illegally carrying 58 avian eggs. During investigation he claimed that the eggs were that of quail, but those were suspected to be parrot eggs. It is illegal to transport parrots or their eggs as 29% of parrot species are declared as endangered species; Priscila F. M and team conducted the barcoding of mtDNA (cytochrome c oxidase subunit I gene [COI] and 16S ribosomal DNA) of embryo sample. They compared the embryonic COI sequences with the BOLD (the Barcode of Life Data System), and the 16S sequences were compared with GenBank sequences. The results were surprising, and on the basis of both the results, they identified that the 57 eggs were of parrots (*Alipiopsitta xanthops*, *Ara ararauna*, and the *Amazona aestival*/*A. ochrocephala*) and 1 was of owl and 1 was an owl.

9 Conclusion

All Bird Barcoding Initiative (ABBI) is a project to explore the biodiversity of birds all over the world. It was established in 2003, and from then thousands of aves species have been identified using my DNA barcoding technique. ABBI seems to be a new hope for identification of novel avian species. There are many relevant applications of COI sequence library submitted on BOLD. Although many studies have been conducted for avian species covering wide geographical area, some areas are better covered than others. BOLD helps to identify these areas where future DNA barcode of avian species should be undertaken. With the advancement of COI library via DNA barcoding projects of avian species, there will be better understanding of different avian realms and taxonomic territories. It also serves as an unbiased taxonomic representation of different avian groups. The advantage is that the DNA barcode sequences deposited in these databases are of high quality and are standardized and therefore have fewer ambiguities, short sequence span, bidirectional sequencing and uniform sequence alignment. It has been observed that the rate of error on BOLD is much lower than on other databases. DNA barcoding in avian species in future will undoubtedly provide more specific species identification,

recognition of cryptic species and tracing the line of avian evolution through different eras and deciphering their causes of divergence. Many studies are already conducted to identify the new species of birds (migratory, sea birds or the birds of Iceland). DNA barcoding also plays a crucial role in forensic investigation of illegal transport of birds or their eggs. Although this became very important tool to identify species of aves, still there are many species of aves to be identified in future.

Acknowledgement The authors would like to acknowledge the University of Tabuk, Tabuk, Saudi Arabia.

The author would also like to thank the Department of Biology, Faculty of Sciences, Saudi Digital Library and University Library for providing the facility for literature survey and collection.

References

AOU (1998) The American Ornithologists' Union. AOU Press, Lawrence, Kansas. Check-list of North American birds (7th edn), 829p

Aliabadian M, Beentjes KK, Roselaar CS (2013) DNA barcoding of Dutch birds. In: Nagy ZT, Backeljau T, De Meyer M, Jordaens K (eds) DNA barcoding: a practical tool for fundamental and applied biodiversity research, vol 365, pp 25–48 (ZooKeys)

Arctander P, Fjeldsa J (1994) Avian tissue collections for DNA analysis. Ibis 136:359360

Avise JC, Walker D (1998) Pleistocene phylogeographic effects on avian populations and the speciation process. Proc R Soc Lond B Biol Sci 265:457–463

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T (1987) Intraspecific phylogeography: the mitochondrial bridge between population genetics and systematics. Annu Rev Ecol Syst 18:489–522

Baker A, Edwards S, Hanken J, Schindel D, Stoeckle M (2005) Barcoding life takes flight: all birds barcoding initiative (ABBI), Report of Inaugural Workshop

Baker AJ, Tavares ES, Elbourne RF (2009) Countering criticisms of single mitochondrial DNA gene barcoding in birds. Mol Ecol Resour 9:257268

Banks RC, Cicero C, Dunn JL, Kratter AW, Ouellet H (2000) Forty-second supplement to the Ornithologists' Union checklist of North American birds. Auk 117:847–858

Banks RC, Cicero C, Dunn JL, Kratter AW, Rasmussen PC (2002) Forty-third supplement to the Ornithologists' Union checklist of North American birds. Auk 119:897–906

Banks RC, Cicero C, Dunn JL, Kratter AW, Rasmussen PC (2003) Forty-fourth supplement to the Ornithologists' Union checklist of North American birds. Auk 120:923–931

Campagna L, St Clair JJH, Lougheed SC, Woods RW, Imberti S, Tubaro PL (2012) Divergence between passerine populations from the Malvinas – Falkland Islands and their continental counterparts: a comparative phylogeographical study. Biol J Linn Soc 106:865–879

Chaves BRN, Chaves AV, Nascimento ACA, Chevitarese J, Vasconcelos MF, Santos FR (2015) Barcoding Neotropical birds: assessing the impact of nonmonophyly in a highly diverse group. Mol Ecol Res 15(4):921–931

Clements JF, Schulenberg TS, Iliff MJ, Roberson D, Fredericks TA, Sullivan BL, Wood CL(2015) The eBird/Clements checklist of birds of the world. http://www.birds.cornell.edu/clementschecklist/download/

Coghlan ML, White NE, Parkinson L, Haile J, Spencer PB, Bunce M (2012) Egg forensics: an appraisal of DNA sequencing to assist in species identification of illegally smuggled eggs. Forensic Sci Int Genet 6(2):268–273

Dasmahapatra KK, Mallet J (2006) DNA barcodes: recent successes and future prospects. Heredity 97(4):254–255

Evans T (2007) DNA damage. NEB expressions 2.1, 1–3
Frézal L, Leblois R (2008) Four years of DNA barcoding: current advances and prospects. Infect Genet Evol 8(5):727–736
Gill FB (2007) Ornithology, 3rd edn. W.H. Freeman, New York
Gill FB, Slikas B (1992) Patterns of mitochondrial DNA divergence in North American crested titmice. Condor 94:20–28
Hajibabaei M, deWaard JR, Ivanova NV, Ratnasingham S, Dooh RT, Kirk SL (2005) Critical factors for assembling a high volume of DNA barcodes. Philos Trans R Soc 360:1959–1967. https://doi.org/10.1098/rstb.1727
Hajibabaei M, Smith AM, Janzen DH (2006) A minimalist barcode can identify a specimen whose DNA is degraded. Mol Ecol Notes 6:959–964
Hebert PDN, Cywinska A, Ball SL, JR DW (2003) Biological identifications through DNA barcodes. Proc R Soc Lond B Biol Sci 270:313–321
Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. PLoS Biol 2(10):e312. https://doi.org/10.1371/journal.pbio.0020312
Hoss M, Jaruga P, Zastawny TH, Dizdaroglu M, Paabo S (1996) DNA damage and DNA sequence retrieval from ancient tissues. Nucleic Acids Res 24:1304–1307
Johnsen A, Rindal E, Ericson PGP, Zuccon D, Kerr KCR, Stoeckle MY (2010) DNA barcoding of Scandinavian birds reveals divergent lineages in trans-Atlantic species. J Ornithol 151:565–578. https://doi.org/10.1007/s10336-009-0490-3
Kerr KCR (2010) A cryptic, intergeneric cytochrome oxidase I pseudogene in tyrant flycatchers (family: Tyrannidae). Genome 53(12):1103–1109
Kerr K, Birks SM, Kalyakin MV, Red'kin YA, Koblik EA, Hebert PDN (2009) Filling the gap – COI barcode resolution in eastern Palearctic birds. Front Zool 6:10
Lijtmaer DA, Kerr KC, Barreira AS, Hebert PD, Tubaro PL (2011) DNA barcode libraries provide insight into continental patterns of avian diversification. PLoS One 6(7):e20744
Lindahl T (1993) Instability and decay of the primary structure of DNA. Nature 362:709–715
Lohman DJ, Ingram KK, Prawiradilaga DM, Winker K, Sheldon FH, Moyle RG (2010) Cryptic genetic diversity in "widespread" Southeast Asian bird species suggests that Philippine avian endemism is gravely underestimated. Biol Conserv 143:1885–1890
Milá B, Tavares ES, Muñoz Saldaña A, Karubian J, Smith TB, Baker AJ (2012) A trans-amazonian screening of mtDNA reveals deep intraspecific divergence in forest birds and suggests a vast underestimation of species diversity. PLoS One 7(7):e40541
Millar CD, Huynen L, Subramanian S, Mohandesan E, Lambert DM (2008) New developments in ancient genomics. Trends Ecol Evol 23:386–393
Moore WS (1995) Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution 49(4):718
Murray BW, McGillivray WB, Barlow JC, Beech RN, Strobeck C (1994) The use of cytochrome *B* sequence variation in estimation of phylogeny in the *Vireonidae*. Condor 96:1037–1054
Newton I (2003) The speciation and biogeography of birds. Academic, London
Nishiumi I, Kim CH (2015) Assessing the potential for reverse colonization among Japanese birds by mining DNA barcode data. J Ornithol 156(S1):325–331
Rosen GE, Smith KF (2010) Summarizing the evidence on the international trade in illegal wildlife. EcoHealth 7:24–32
Seutin GW, Bradley N, Boag PT (1991) Preservation of avian blood and tissue samples for DNA analyses. Can J Zool 69:82–90
Stoeckle M (2005) Barcoding life takes flight: all birds barcoding initiative (ABBI), needs and resources statement
Stoeckle MY, Kerr KCR, Zuo Z (2012) Frequency matrix approach demonstrates high sequence quality in avian BARCODEs and highlights cryptic Pseudogenes. PLoS One 7(8):e43992
Tavares ES, Baker AJ (2008) Single mitochondrial gene barcodes reliably identify sister-species in diverse clades of birds. BMC Evol Biol 8(1):81

Tavares ES, Gonçalves P, Miyaki CY, Baker AJ, Schierwater B (2011) DNA barcode detects high genetic structure within Neotropical bird species. PLoS One 6(12):e28543

Woese CR (2000) Interpreting the universal phylogenetic tree. Proc Natl Acad Sci U S A 97:8392–8396

Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A 74:5088–5090

Yoo HY, Eah J, Kim JS, Young JK, Min M, Peak WK (2006) DNA barcoding Korean birds. Mol Cell 22:323–327

Zimmermann J, Hajibabaei M, Blackburn D (2008) DNA damage in preserved specimens and tissue samples: a molecular assessment. Front Zool 5:18

Molecular Characterization of Ruminant Mammals Using DNA Barcodes

Muniyandi Nagarajan, Koodali Nimisha, and Subhash Thomas

Abstract Ruminant mammals are widely distributed across the world and distinguished from other mammals by the presence of four-chambered stomach. Most of the ruminants are wild, while a few are domestic which contribute significantly to the agricultural economy in the form of livestock resources. Characterization of livestock breeds and the exact identification of wild ruminant species are imperative for developing improved breeds and wildlife conservation, respectively. Though, taxonomists determine breeds and species based on morphological traits, which is nugatory in the case of cryptic species or when unrelated species exhibit similar morphological traits. However, the emerging DNA-based technique has overcome the challenges and limitations faced by the conventional methods. DNA barcoding, specifically, the discovery of mitochondrial cytochrome *c* oxidase subunit I (COI) gene as a standard DNA barcode region for animals, has transfigured the realm of molecular systematics by providing a platform to expeditiously find novel lineages and elucidate ruminant phylogeny. Despite this, DNA barcoding has huge applications including detection of adulteration and mislabeling of bush meat and checking wildlife poaching and animal trafficking. This chapter provides an overview of ruminant mammals and the usefulness of COI gene in the identification of ruminant species.

Keywords Ruminant · Mammal · COI gene · Phylogeny

1 A Glimpse on Ruminant Mammals

Mammals can be basically defined as a group of vertebrate species possessing unique features such as presence of mammary glands, a single bone in the lower jaw, and neocortex of the forebrain, which are found completely lacking in other vertebrates

M. Nagarajan (✉) · K. Nimisha · S. Thomas
Department of Genomic Science, School of Biological Sciences, Central University of Kerala, Kasargod, Kerala, India
e-mail: nagarajan@cukerala.ac.in

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_20

(Kemp 2004). Around the globe, a total of 5416 species of mammals have been reported so far and are placed under 154 families and 29 orders (Wilson and Reeder 2005). Of these, India possesses a total of 423 species, which makes 7.81% of the global mammalian species representing 48 families and 14 orders (Sharma et al. 2013). Among the various mammalian orders, the suborder Ruminantia (Order, Artiodactlya) holds a special position as most of them are livestock and thus are contributing indispensably to the world's economic growth. Ruminants are known for their complex digestive system consisting of "rumen," a chamber specialized in microbial fermentation and regurgitation of food for its further break down and digestion (Fig. 1). These large terrestrial herbivores enjoy ubiquitous distribution and are native to almost all continents like Eurasia, Africa, North America, and South America except Australia and Antarctica (Hassanin and Douzery 2003; Fernández and Vrba 2005; Hackmann and Spain 2010). As of now there are around 200 extant ruminant species of which the widely domesticated species include cattle, buffalo, goat, sheep, mithun, reindeer, yak, and camel. The suborder Ruminantia comprises of six extant families such as Tragulidae, Moschidae, Bovidae, Giraffidae, Antilocapridae and Cervidae. Except Tragulidae, the latter five extant families, belonging to the infraorder Pecora, are popularly known as the true ruminants (Kulemzina et al. 2011). Among these families, the Bovidae family is the most

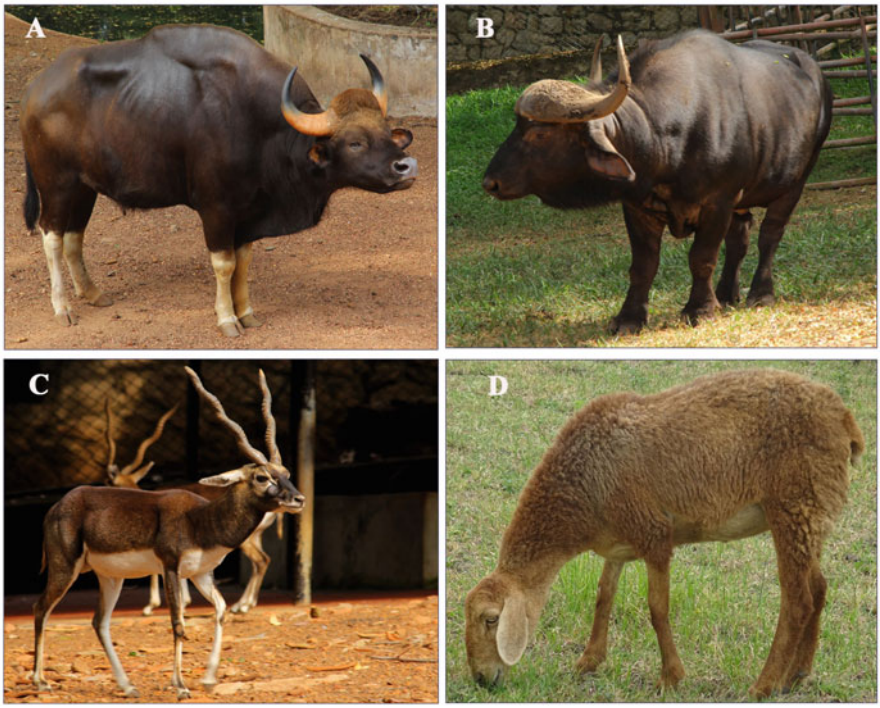


Fig. 1 The ruminant mammals. (**a**) Gaur (*Bos gaurus*). (**b**) African buffalo (*Syncerus caffer*). (**c**) Blackbuck (*Antilope cervicapra*). (**d**) Sheep (*Ovis aries*)

diversified one (Kraus and Miyamoto 1991; Heywood 2010), and majority of the livestock species are placed under this family. As per the 19th livestock census (2012) reports, total livestock population of India is 512.05 million which includes bovine (299.9 million), sheep (65.06 million), and goats (135.17 million).

First ruminants were believed to have evolved around 50 million years BC in the middle of Eocene epoch at certain parts of Northern America and Eurasia (Gentry 1994; Hackmann and Spain 2010). Since prehistoric times, domesticated ruminants were closely associated with human population. Throughout the world, livestock are valued for their economic output such as meat (as dietary protein source), dairy products, fiber (wool, hair, hooves, and skin from sheep, goats), fertilizers (from animal bones), muscle power (for transportation and plowing), and land management. Global demand for livestock and livestock products is increasing rapidly with expanding human population especially in developing nations (Thornton 2010).

To meet with the increasing demand of livestock and livestock products, extensive crossbreeding and selective breeding have been practiced in many of the ruminant species. These practices often lead to a decline in the indigenous varieties of breeds mostly because they are not as productive as the crossbreeds or may be their profitability under present market conditions are very low. Therefore, many of the local breeds are facing extinction. Also, the genetic variability in the available gene pool is getting eroded continuously. Under such circumstances, a proper assessment of existing breeds becomes essential for proper conservation and genetic improvement of the germplasm. Molecular approaches provide a wide opportunity for breed assessment, as there are numerous tools like DNA markers, which can greatly contribute to genetic variability evaluation within the breeds as well as between the breeds.

2 Trends in Molecular Phylogeny of Ruminant Mammals

Ruminants, being very diverse in their ecology, behavior, physiology, and biogeographical distribution, are subjects of interest to many biologists. The diversity among them can be systematically studied using phylogeny. In general phylogeny relies on variations at morphological or molecular levels to deduce the evolutionary relationship between taxonomic groups. Molecular phylogeny utilizes inheritable structural or functional biomolecular character sets for constructing phylogenetic tree. A well-resolved phylogenetic tree can describe species relationship, population history, and dynamics of evolution.

The construction of phylogenetic tree had always remained a challenge for taxonomists perhaps, due to morphological convergence of unrelated taxa or difficulty in discriminating cryptic species. The first molecular approach towards the identification of livestock diversity was based on protein polymorphisms (Avise 1994). A large number of studies have been carried out in livestock by characterizing allozyme system and blood group. Since the level of polymorphism exhibited by

blood group and allozyme system is low, the suitability of protein markers was found inadequate in identifying animals at species level. In contrast, DNA polymorphisms have shown greater efficiency in precisely detecting species. The polymorphic DNA markers used in characterizing livestock generally include RFLP (restriction fragment length polymorphism), AFLP (amplified fragment length polymorphism), RAPD (random amplified polymorphic DNA), microsatellites, and mitochondrial and Y chromosomal markers. Numerous studies have been conducted on ruminants (goat, sheep, cattle, buffalo, mithun, and camel) using RFLP, AFLP, and RAPD markers, particularly, during the 1990s (Xiang-long et al. 1997; Ajmone-Marsan et al. 2001; de Mello Klocker Vasconcellos et al. 2003; Saifi et al. 2004; Sodhi et al. 2006; Khaldi et al. 2010; Mahrous and Ramadan 2011). Subsequent studies have characterized livestock that include small ruminants (goat and sheep) as well as large herbivores (buffalo, cattle, mithun, yak, and camel) using microsatellite markers (Santos-Silva et al. 2008; Bozzi et al. 2009; Kumar et al. 2006; Nagarajan et al. 2009). As molecular tools for determining species developed, great effort was put forth to characterize variable regions of mitochondria such as D-loop region, cytochrome *b* gene, and COI gene (Lai et al. 2006; Kumar et al. 2007; Sanches et al. 2012; Nagarajan et al. 2015). However, COI gene stands out as the best molecular toolkit for carrying out genetic diversity studies since it is widely accepted as the universal DNA barcode region (exclusively for animals) due to high evolutionary rate. The COI gene has opened up new avenues for identification ruminant species and understanding the genetic relationship between various ruminants.

3 Emergence of DNA Barcoding as a Novel Tool

In 2003, it was proposed that mitochondrial DNA sequences of cytochrome c oxidase subunit I gene can be used as "barcodes" for the bio-identification of species globally (Hebert et al. 2003a, b). This novel technology uses the variations arising in the short DNA sequences as labels for species. Initially, utility of this technique was established in Lepidopterans (Hebert et al. 2004a; Janzen et al. 2005; Hajibabaei et al. 2006); later it became widespread around the animal world which included invertebrate species (Barrett and Hebert 2005; Wang et al. 2011), fishes (Ward et al. 2005), amphibians (Vences et al. 2005), reptiles (Vences et al. 2012; Nagy et al. 2012), birds (Hebert et al. 2004b), and mammals (Cai et al. 2011; Yan et al. 2013). The 651bp fragment of mitochondrial DNA COI gene has been recognized as "standard barcode" because of its peculiarities such as short size and availability of universal primers.

The mitochondrial DNA COI gene, which is responsible for carrying out oxidative phosphorylation, is highly conserved across the species (Waugh 2007). The nucleotide variation in the gene is efficient enough to distinguish animals at interspecific level. However, the intraspecific variation has been reported to be less than 10%. In addition to this, COI gene rarely possesses any insertions or deletions (Waugh 2007). These properties make it COI gene ideal for DNA barcoding, a

highly promising tool in disclosing the identity of cryptic species (Hebert et al. 2004a), distinguishing species at their juvenile stages (Valdez-Moreno et al. 2010) and in protecting endangered species (Elmeer et al. 2012). It is achieved by generating barcode data of several species and depositing it on the public domain, further using these sequences for identification of unknown specimens. There are several international barcoding projects focused on developing barcode sequence libraries for specific targeted species like "Mammalia Barcode of Life" campaign, a project that aims to build a comprehensive reference library of DNA barcodes for the global mammal fauna.

The COI gene sequences have been widely used to identify ruminant species in different studies. Cai et al. (2011) analyzed sequences of 223 individuals of Bovidae representing 18 species. The result showed that, except two species, all other species studied possessed unique COI sequences. Similarly, Arif et al. (2012) performed a comparative study to characterize 12 members of the family Bovidae using different mitochondrial markers including COI gene. In the phylogenetic tree, the genus *Bison* and *Bos* showed close relationship and supported the previous studies suggesting two lineages of the tribe Bovini-buffalo (*Bubalus*) and cattle (*Bos*). The study showed a main split of the 12 members of the Bovidae into bovine clade and non-bovine clade. In another study, COI barcodes helped in deciphering the identity of Tanzanian Bovidae species (Bitanyi et al. 2011). A 470 bp region of the COI gene was tested in 95 specimens representing 20 species of antelopes, wild buffalo, and domesticated species of Bovidae (*Bos taurus*, *B. indicus*, *Ovis aries*, and *Capra hircus*). In this particular study, even 50 bp COI sequence was proved to be effective in discriminating species as this region was found to be highly variable among the species, and also this variability was evenly distributed along the gene fragment.

Yan et al. (2013) have reported that COI gene can be used as potential tool for the identification of Bovidae (*Ovis aries*, *Capra hircus*, *Bos grunniens*, *B. taurus*, *Bubalus bubalis*, *Saiga tatarica*, *Procapra gutturosa*) and Cervidae (*Cervus elaphus*, *C. nippon*, and *Elaphurus davidianus*) species even from traces of animal parts such as horns. The study was performed by analyzing 223 mitochondrial COI sequences, which included sequences generated from 47 specimens of animal horn derived from 10 known species and 176 COI sequences retrieved from GenBank. The results showed less than 1.4% variation within species whereas above 2.0% variation between species; thus it was possible to identify and discriminate the species using samples of animal horn. Bondoc and Cerbito (2013) have shown the effectiveness of DNA barcoding in establishing the genetic relationship within sheep and goat breeds in the Philippines. Likewise, Ali et al. (2016) used COI gene sequences to distinguish the native goat breeds of Pakistan from the exotic goat breeds.

DNA barcoding (COI gene) has also served as a tool in forensic identification of specimens. A case study from South Africa utilized COI gene as genetic marker to determine the species identity of unknown samples of meat obtained from three different sources in different forms: dried sample, frozen sample, and carcasses alleged to be from common reedbuck (*Redunca arundinum*). The analysis showed that the dried and frozen samples of meat belonged to that of domestic cattle (*Bos taurus*) and the carcasses was confirmed to be of common reedbuck as the COI gene

sequences of the specimen matched with that of reference sequences in the database (Dalton and Kotze 2011). It was very difficult to identify the Chinese deer as they pronounce morphological similarities until Cai et al. (2015) reported unique barcodes for 21 species of the family Cervidae.

The use of COI maker has been extended to forensic identification of mislabeled or misbranded game meat or bush meat products. Bush meat or game meat is commonly referred to the meat of undomesticated animal's especially wild animals, which are hunted for food. Many of the ruminant species are exploited by hunting. Most common among them are antelopes like kudu, eland, deer, and Cape buffalo. Though game meat hunting is legal in many nations like South Africa and Namibia, meat mislabeling is a common concern prevailing in these areas. Several studies have been reported with regard to meat mislabeling; however identification of game meat becomes difficult due to conversion of poached game meat into unrecognizable form by removing body portions or by processing. A DNA barcoding study based on COI gene was conducted in 146 samples (14 beef and 132 game labels) of meat products to check the authenticity of commercial labels found in the local market. About 76.5% mislabeling was detected with the game meat samples, which showed substitution of the game meat with the meat from domestic cattle, pig, lamb, and horse. Additionally, meat of less common species such as giraffe, waterbuck, bushbuck, duiker, and mountain zebra were also found to be substituted with game meat (D'Amato et al. 2013). A similar study by Quinto et al. (2016) reported the game meat mislabeling in US market. A 658bp region of COI gene was used as a maker for analyzing 54 game meat samples representing variety of species. The results showed 10 out of 54 samples as mislabeled. Two products labeled as bison and yak were found to be of domestic cattle. The overall rate of mislabeling in this study was found to be 18.5%. Similarly, Kane and Hellberg (2016) analyzed the meat purchased from three different sources: online sources, supermarkets, and local butcher shops. The study showed higher rate of mislabeling (38%) in meat products bought from online dealers compared to that of butcher shops (18%) and supermarket (5.8%).

Syakalima et al. (2016) robustly identified nine ruminant species (puku, eland, impala, Kafue lechwe, bushbuck, waterbuck, buffalo, wildebeest, and sable) from poached game meat of wild ruminant species by comparing the COI gene sequences with that of reference database (BOLD). Similarly, illegal hunting of endangered Pampas deer has been reported from Brazil using COI (Sanches et al. 2012). Mbugua (2014) carried out DNA barcoding studies in 99 unknown meat samples collected from different parts of Kenya along with 5 wild species, namely, impala (*Aepyceros melampus*), waterbuck (*Kobus ellipsiprymnus*), African buffalo (*Syncerus caffer*), black rhino (*Diceros bicornis*), and elephant (*Loxodonta cyclotis*) in order to identify the prevalence of bush meat and meat product substitution. COI gene successfully revealed the identity of the species from the tissue samples, which showed the meat substitutions in the samples analyzed.

Understanding the relationship between hosts, parasites, and their intermediate vectors is important in establishing the epidemiology of zoonotic diseases. Since the collection and identification of parasites are difficult using conventional methods,

DNA barcoding approach can be utilized for identifying the parasites as well as the diversity of them. Efficient protocols for DNA barcoding have been reported for identifying most of the parasites affecting ruminants including filarioid nematodes, ticks, mosquitoes, and sarcocystis (Cywinska et al. 2006; Ferri et al. 2009; Gjerde 2013; Zhang and Zhang 2014). DNA barcoding through scat analysis is a convenient method for identifying parasites that bypasses the euthanization of animals. Being a noninvasive technique, it also extends the study to be conducted in live animals, unlike in conventional study where study could be performed only in diseased or dead animals.

4 A Composite Phylogenetic Analysis of Ruminants Using COI Gene

Even though previous studies have proven the efficiency of COI barcodes in delineating species, a comprehensive phylogenetic analysis has not performed on ruminants using COI gene sequences, yet. Thus to represent the interrelationship between different ruminant species and to reaffirm the efficiency of COI marker in identification, COI gene sequences of 22 ruminant species were retrieved from the GenBank and analyzed (Table 1). The sequences were aligned using ClustalW implemented in MEGA (Tamura et al. 2011) and, subsequently, truncated to 523 bp to make it equal in size for further analysis. The maximum parsimony phylogenetic tree was constructed using 100 COI gene sequences with a bootstrap value of 1000 using the software MEGA. It formed three distinct major clades with a few internal subclades (Fig. 2). Among the three, the first clade consists of sheep (domestic sheep, Dall sheep, and bighorn sheep), goat (Siberian ibex, domestic goat), Himalayan blue sheep, African buffalo, and domestic buffalo. The second clade forms a group exclusively of *Bos* and *Bison* species, whereas the third clade comprises members of the family Camelidae.

The maximum parsimony tree topology clearly showed the close evolutionary relationship between sheep and goat with buffalo though they are classified under different subfamily—Caprinae and Bovinae, respectively—which was comparable to previously published reports (Fernández and Vrba 2005). The evolutionary relationship and systematic classification of Bovidae family remain as a dispute and different classification systems exist for bovids based on the phyletic relationship. Morphological and molecular examination has suggested considering the Bovidae family as a monophyletic group, where all the descendants evolved from a common ancestor forms a clade. However, establishment of monophyly has been weakly supported by previous studies. In contradiction with this statement, paraphyletic grouping (where most of the descendants of a common evolutionary ancestor form a group excluding a few descendants which form a separate group) of family Bovidae has been described in several literatures (Gatesy et al. 1992). In concordance with the study of Gatesy et al. (1992), our phylogenetic tree analysis

Table 1 Details of COI sequences used for phylogenetic tree construction

S. No	Species	No. of sequences	GenBank ID
1	*Bison bison*	6	JF443190-JF443195
2	*Bison bonasus*	2	EU623450, JF444283
3	*Bos frontalis*	1	HQ269429
4	*Bos gaurus*	1	KF808255
5	*Bos grunniens*	8	HQ269432, HQ269433, HQ269462- HQ269467
6	*Bos indicus*	10	KF952275 - KF952279, KF952281- KF952285
7	*Bos primigenius*	2	JQ735452, JQ735453
8	*Bos taurus*	10	HQ860420, JF700140, JF700141, JX567086, KF771228, KF799986, KU947025- KU947028
9	*Bubalus bubalis*	10	KU932060, KU932061, KU932067, KU932072 KU932073, KU932075, KU932096, KU932100, KU932106, KU932110
10	*Capra hircus*	10	HQ269454, JN850776, JN850777, JQ735456, KC679016- KC679018, KF317903, KF317905, KT750041
11	*Capra sibirica*	1	KU527896
12	*Ovis aries*	10	JN245995, JN850771- JN850773, JQ735465, JX567087, KF317902, KF317907, KT750038, KT750039
13	*Ovis canadensis*	3	JF443356- JF443358
14	*Ovis dalli*	2	JF443359, JF443360
15	*Pseudois nayaur*	4	HQ269457- HQ269459, KJ862174
16	*Syncerus caffer*	2	JN082178, KF482455
17	*Camelus dromedarius*	10	JN632608, AB753111- AB753115, AB753117- AB753120
18	*Camelus bactrianus*	5	AP003423, EF507798- EF507801
19	*Lama pacos*	2	AJ566364, DQ534055
20	*Lama glama*	1	DQ534054
21	*Lama guanicoe*	1	DQ534053
22	*Vicugna vicugna*	1	DQ534056

quite apparently has represented two distinct lineages of the subfamily Bovinae, one which forms a distinct clade (clade II) comprising bos and bison groups and the latter consisting of African and Indian buffalo that forms a different group with members of Caprinae family (Arif et al. 2012).

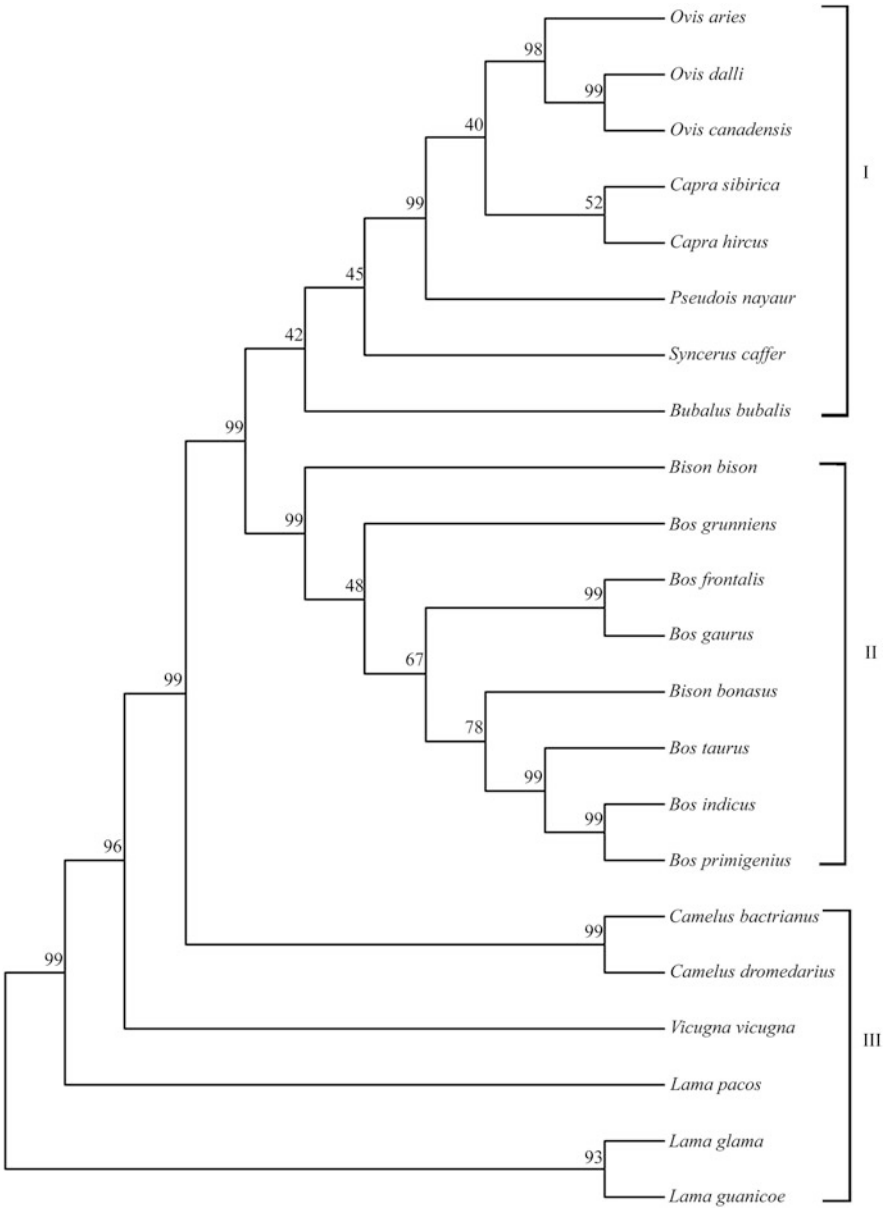


Fig. 2 Maximum parsimony phylogenetic tree of selected ruminant mammals

It is believed that cattle species have evolved from *Bos primigenius*. Studies have recorded that the ancestral species has already become extinct long back. The maximum parsimony tree showed that *Bos indicus* genetically is closer to the extinct

Bos primigenius than *B. taurus*. It also showed the close genetic relationship between *Bos* and *Bison* genera. The close relationship between these two genera is well supported by the evidences obtained from paleontological and reproductive data (Arif et al. 2012). Fernández and Vrba (2005) have even suggested combining the genus *Bos* and *Bison* into a single genus. Several studies, based on microsatellites, mtDNA, and nuclear DNA, including our analysis of CO I gene, have demonstrated that yaks (*Bos grunniens*) are genetically closer to American bison (*Bison bison*) and distant from *Bison bonasus* (Zhengchao et al. 1998; Tu et al. 2002; Hassanin and Ropiquet 2004; Lai et al. 2006; Xie et al. 2010; Qiu et al. 2012; Bai 2015). The third clade comprising members of Camelidae appears as a distinct group from the bovids. Also, a clear distinction can be spotted between the Asian, African, and New World camels. The Ilama (*Lama glama*) and guanaco (*Lama guanicoe*) appear to be diverging from alpaca (*Lama pacos*) and vicugna (*Vicugna vicugna*) as reported in the previous studies using cytochrome *b* sequence (Stanley et al. 1994).

The phylogenetic analysis is necessary for understanding the evolution of diverse taxa and making most probable prediction about cryptic or poorly known species. COI gene has clearly distinguished different ruminant species, and thus it can serve as an efficient tool for the identification of ruminants as well as systematic organization of the species into various taxa.

5 Conclusion

DNA barcoding renders a potent identification system for ruminant species. Besides this, DNA barcodes have paved the way for labeling of livestock products especially meat and milk, which authenticate the composition of livestock products thereby protecting the consumer from being misguided and cheated by illegal substitutions. Among the various molecular markers used in molecular systematics, COI gene has received much attention among the molecular taxonomists as they are effective in discriminating congeneric taxa. Since most of the ruminants are wild animals, DNA barcoding serves as an effective method to develop strategies for conserving ruminant species facing extinction. On the other hand, it has also contributed in the identification of parasites that cause ailments in ruminant livestock. Parasites are identified mostly by analyzing feces of ruminants, which further provides a noninvasive alternative way of disease identification without disturbing the animal. A much more recent approach that integrates DNA barcoding with next-generation sequencing technology has even made it possible to analyze the rumen contents of large ruminants with more precision generating sequences for multiple samples at once. We envision that in near future, ruminant research will be greatly benefitted by the advances in DNA barcoding technique, which can serve as a key for the sustainable management of endangered wild species as well as indigenous livestock breeds.

Acknowledgments The senior author is thankful to the Science and Engineering Research Board (SERB), Department of Science and Technology, Government of India, New Delhi for the financial

Assistance (SR/S0/AS-84/2012). ST is thankful to the Kerala State Council for Science, Technology and Environment (KSCSTE) for the support in the form of research fellowship.

References

Ajmone-Marsan P, Negrini R, Crepaldi P et al (2001) Assessing genetic diversity in Italian goat populations using AFLP markers. Anim Genet 32:281–288

Ali A, Rehman A, William K (2016) Phylogenetic analysis of *Capra hircus* commonly found goat breeds of Pakistan using DNA barcode. J Bioresour Manage 3(1)

Arif IA, Bakir MA, Khan HA (2012) Inferring the phylogeny of bovidae using mitochondrial DNA sequences: resolving power of individual genes relative to complete genomes. Evol Bioinformatics Online 8:139–150. https://doi.org/10.4137/EBO.S8897

Avise JC (1994) History of molecular phylogenetics. In: Avise JC (ed) Molecular markers, natural history and evolution. Springer, pp 16–43

Bai JL (2015) Phylotaxonomic position of Tianzhu white yak (*Poephagus grunniens*) based on nucleotide sequences of multiple subunits of cytochrome *c* oxidase. J Appl Anim Res 43:431–438. https://doi.org/10.1080/09712119.2014.980416

Barrett RDH, Hebert PDN (2005) Identifying spiders through DNA barcodes. Can J Zool 83:481–491. https://doi.org/10.1139/z05-024

Bitanyi S, Bjørnstad G, Ernest EM et al (2011) Species identification of Tanzanian antelopes using DNA barcoding. Mol Ecol Resour 11:442–449. https://doi.org/10.1111/j.1755-0998.2011.02980.x

Bondoc OL, Cerbito WA (2013) Genetic diversity and relationships of domestic goat and sheep breeds (Artiodactyla: Bovidae: Caprinae) in the Philippines based on DNA barcodes. J Vet Sci 50(2):64–74

Bozzi R, Degl'Innocenti P, Rivera Diaz P et al (2009) Genetic characterization and breed assignment in five Italian sheep breeds using microsatellite markers. Small Rumin Res 85:50–57. https://doi.org/10.1016/j.smallrumres.2009.07.005

Cai Y, Zhang L, Shen F et al (2011) DNA barcoding of 18 species of Bovidae. Chin Sci Bull 56:164–168. https://doi.org/10.1007/s11434-010-4302-1

Cai Y, Zhang L, Wang Y et al (2015) Identification of deer species (Cervidae, Cetartiodactyla) in China using mitochondrial cytochrome c oxidase subunit I (mtDNA COI). Mitochondrial DNA 1736:1–4. https://doi.org/10.3109/19401736.2014.1003919

Cywinska A, Hunter FF, Hebert PDN (2006) Identifying Canadian mosquito species through DNA barcodes. Med Vet Entomol 20:413–424. https://doi.org/10.1111/j.1365-2915.2006.00653.x

D'Amato ME, Alechine E, Cloete KW et al (2013) Where is the game? Wild meat products authentication in South Africa: a case study. Investig Genet 4:6. https://doi.org/10.1186/2041-2223-4-6

Dalton DL, Kotze A (2011) DNA barcoding as a tool for species identification in three forensic wildlife cases in South Africa. Forensic Sci Int 207:e51–e54. https://doi.org/10.1016/j.forsciint.2010.12.017

de Mello Klocker Vasconcellos LP, Tambasco-Talhari D, Pozzi Pereira A et al (2003) Genetic characterization of Aberdeen Angus cattle using molecular markers. Genet Mol Biol 26:133–137. https://doi.org/10.1590/S1415-47572003000200005

Elmeer K, Almalki A, Mohran KA et al (2012) DNA barcoding of *Oryx leucoryx* using the mitochondrial cytochrome C oxidase gene. Genet Mol Res 11(1):539–547

Fernández MH, Vrba ES (2005) A complete estimate of the phylogenetic relationships in Ruminantia: a dated species-level supertree of the extant ruminants. Biol Rev 80:269–302. https://doi.org/10.1017/S1464793104006670

Ferri E, Barbuto M, Bain O et al (2009) Integrated taxonomy: traditional approach and DNA barcoding for the identification of filarioid worms and related parasites (Nematoda). Front Zool 6(1). https://doi.org/10.1186/1742-9994-6-1

Gatesy J, Yelon D, Desall R, Vrba ES (1992) Phylogeny of the Bovidae (Artiodactyla, Mammalia), based on mitochondrial ribosomal DNA sequences. Mol Biol Evol 9(3):433–446

Gentry AW (1994) The Miocene differentiation of old world Pecora (Mammalia). Hist Biol 7:115–158. https://doi.org/10.1080/10292389409380449

Gjerde B (2013) Phylogenetic relationships among Sarcocystis species in cervids, cattle and sheep inferred from the mitochondrial cytochrome c oxidase subunit I gene. Int J Parasitol 43:579–591. https://doi.org/10.1016/j.ijpara.2013.02.004

Hackmann TJ, Spain JN (2010) Invited review: ruminant ecology and evolution: perspectives useful to ruminant livestock research and production. J Dairy Sci 93:1320–1334. https://doi.org/10.3168/jds.2009-2071

Hajibabaei M, Janzen DH, Burns JM et al (2006) DNA barcodes distinguish species of tropical Lepidoptera. Proc Natl Acad Sci U S A 103:968–971. https://doi.org/10.1073/pnas.0510466103

Hassanin A, Douzery EJP (2003) Molecular and morphological phylogenies of ruminantia and the alternative position of the moschidae. Syst Biol 52:206–228

Hassanin A, Ropiquet A (2004) Molecular phylogeny of the tribe Bovini (Bovidae, Bovinae) and the taxonomic status of the Kouprey, *Bos sauveli* Urbain 1937. Mol Phylogenet Evol 33:896–907. https://doi.org/10.1016/j.ympev.2004.08.009

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003a) Biological identifications through DNA barcodes. Proc Biol Sci 270:313–321. https://doi.org/10.1098/rspb.2002.2218

Hebert PDN, Ratnasingham S, Waard J (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc R Soc Lond B 270: S96–S99. https://doi.org/10.1098/rsbl.2003.0025

Hebert PDN, Penton EH, Burns JM et al (2004a) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc Natl Acad Sci U S A 101:14812–14817

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004b) Identification of birds through DNA barcodes. PLoS Biol 2(10):e312

Heywood JJN (2010) Explaining patterns in modern ruminant diversity: contingency or constraint? Biol J Linn Soc 99:657–672. https://doi.org/10.1111/j.1095-8312.2010.01436.x

Janzen DH, Hajibabaei M, Burns JM et al (2005) Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. Philos Trans R Soc Lond Ser B Biol Sci 360:1835–1845. https://doi.org/10.1098/rstb.2005.1715

Kane DE, Hellberg RS (2016) Identification of species in ground meat products sold on the U.S. commercial market using DNA-based methods. Food Control 59:158–163. https://doi.org/10.1016/j.foodcont.2015.05.020

Kemp TS (2004) The origin and evolution of mammals. Oxford University Press, New York

Khaldi Z, Rekik B, Haddad B, et al (2010) Genetic characterization of three ovine breeds in Tunisia using randomly amplified polymorphic DNA markers

Kraus F, Miyamoto MM (1991) Rapid cladogenesis among the pecoran ruminants: evidence from mitochondrial DNA sequences. Syst Zool 40:117–130

Kulemzina AI, Yang F, Trifonov VA et al (2011) Chromosome painting in Tragulidae facilitates the reconstruction of Ruminantia ancestral karyotype. Chromosom Res 19:531–539. https://doi.org/10.1007/s10577-011-9201-z

Kumar S, Gupta J, Kumar N, Dikshit K, Navani N, Jain P, Nagarajan M (2006) Genetic variation and relationships among eight Indian riverine buffalo breeds. Mol Ecol 15:593–600

Kumar S, Nagarajan M, Sandhu JS, Kumar N, Behl V, Nishanth G (2007) Mitochondrial DNA analyses of Indian water buffalo support a distinct genetic origin of river and swamp buffalo. Anim Genet 38:227–232

Lai SJ, Liu YP, Liu YX, Li XW, Yao YG (2006) Genetic diversity and origin of Chinese cattle revealed by mtDNA D-loop sequence variation. Mol Phylogenet Evol 38:146–154

Mahrous K, Ramadan H (2011) Genetic variations between camel breeds using microsatellite markers and RAPD techniques. J Appl Biosci 39:2626–2634

Mbugua D (2014) Use of cytochrome oxidase 1 gene region: a molecular tool for the domestic and wildlife industry in Kenya. Sci J Anim Sci 3(3):87–94

Nagarajan M, Kumar N, Nishanth G et al (2009) Microsatellite markers of water buffalo, *Bubalus bubalis*- development, characterisation and linkage disequilibrium studies. BMC Genet 10:68. https://doi.org/10.1186/1471-2156-10-68

Nagarajan M, Nimisha K, Kumar S (2015) Mitochondrial DNA variability of domestic river buffalo (*Bubalus bubalis*) populations: genetic evidence for domestication of river buffalo in Indian subcontinent. Genome Biol Evol 7:1252–1259. https://doi.org/10.1093/gbe/evv067

Nagy ZT, Sonet G, Glaw F, Vences M (2012) First large-scale DNA barcoding assessment of reptiles in the biodiversity hotspot of Madagascar, based on newly designed COI primers. PLoS One 7:e34506

Qiu Q, Zhang G, Ma T et al (2012) The yak genome and adaptation to life at high altitude. Nat Genet 44:946–949

Quinto CA, Tinoco R, Hellberg RS (2016) DNA barcoding reveals mislabeling of game meat species on the U.S. commercial market. Food Control 59:386–392. https://doi.org/10.1016/j.foodcont.2015.05.043

Saifi HW, Bhushan B, Kumar S et al (2004) Genetic identity between Bhadawari and Murrah breeds of Indian buffaloes (*Bubalus bubalis*) using RAPD-PCR. Asian-Australas J Anim Sci 17:603–607. https://doi.org/10.5713/ajas.2004.603

Sanches A, Tokumoto PM, Peres WAM et al (2012) Illegal hunting cases detected with molecular forensics in Brazil. Investig Genet 3:17. https://doi.org/10.1186/2041-2223-3-17

Santos-Silva F, Ivo RS, Sousa MCO et al (2008) Assessing genetic diversity and differentiation in Portuguese coarse-wool sheep breeds with microsatellite markers. Small Rumin Res 78:32–40. https://doi.org/10.1016/j.smallrumres.2008.04.006

Sharma G, Kamalakannan M, Venkataraman K (2013) A checklist of mammals of India with their distribution and conservation status. Director Zoological Survey of India

Sodhi M, Mukesh M, Anand A et al (2006) Assessment of genetic variability in two North Indian Buffalo breeds using random amplified polymorphic DNA (RAPD) markers. Asian-Aust J Anim Sci 19:1234–1239. https://doi.org/10.5713/ajas.2006.1234

Stanley HF, Kadwell M, Wheeler JC (1994) Molecular evolution of the family Camelidae: a mitochondrial DNA study. Proc Biol Sci 256:1–6. https://doi.org/10.1098/rspb.1994.0041

Syakalima M, Munyeme M, Yasuda J (2016) Cytochrome c oxidase sequences of Zambian wildlife helps to identify species of origin of meat. doi: https://doi.org/10.1155/2016/1808912

Tamura K, Peterson D, Peterson N et al (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739. https://doi.org/10.1093/molbev/msr121

Thornton PK (2010) Livestock production: recent trends, future prospects. Philos Trans R Soc Lond Ser B Biol Sci 365:2853–2867. https://doi.org/10.1098/rstb.2010.0134

Tu Z-C, Qiu H, Zhang Y-P (2002) Polymorphism in mitochondrial DNA (mtDNA) of yak (Bos grunniens). Biochem Genet 40:187–193

Valdez-Moreno M, Vásquez-Yeomans L, Elías-Gutiérrez M et al (2010) Using DNA barcodes to connect adults and early life stages of marine fishes from the Yucatan Peninsula, Mexico: potential in fisheries management. Mar Freshw Res 61:655–671

Vences M, Thomas M, Bonett RM, Vieites DR (2005) Deciphering amphibian diversity through DNA barcoding: chances and challenges. Philos Trans R Soc B Biol Sci 360:1859–1868. https://doi.org/10.1098/rstb.2005.1717

Vences M, Nagy ZT, Sonet G, Verheyen E (2012) DNA barcoding amphibians and reptiles. Methods Mol Biol 858:79–107. https://doi.org/10.1007/978-1-61779-591-6_5

Wang JJ-F, Jiang L-Y, Qiao G-X (2011) Use of a mitochondrial COI sequence to identify species of the subtribe Aphidina (Hemiptera, Aphididae). Zookeys 122:1–17. https://doi.org/10.3897/zookeys.122.1256

Ward RD, Zemlak TS, Innes BH et al (2005) DNA barcoding Australia's fish species. Philos Trans R Soc Lond Ser B Biol Sci 360:1847–1857. https://doi.org/10.1098/rstb.2005.1716

Waugh J (2007) DNA barcoding in animal species: progress, potential and pitfalls. BioEssays 29:188–197. https://doi.org/10.1002/bies.20529

Wilson DE, Reeder DAM (2005) Mammal species of the world: a taxonomic and geographic reference. Johns Hopkins University Press

Xiang-long L, Zhang Y, Shen-gou C et al (1997) Study on the mtDNA RFLP of goat breeds. Zool Res 18:421–428

Xie Y, Li Y, Zhao X, et al (2010) Origins of the Chinese Yak: evidence from maternal and paternal inheritance. In: 2010 4th international conference on bioinformatics and biomedical engineering, pp 1–5

Yan D, Luo JY, Han YM et al (2013) Forensic DNA barcoding and bio-response studies of animal horn products used in traditional medicine. PLoS One 8:e55854

Zhang RL, Zhang B (2014) Prospects of using DNA barcoding for species identification and evaluation of the accuracy of sequence databases for ticks (Acari: Ixodida). Ticks Tick Borne Dis 5:352–358. https://doi.org/10.1016/j.ttbdis.2014.01.001

Zhengchao T, Yaping Z, Huai Q (1998) Mitochondrial DNA polymorphism and genetic diversity in Chinese yaks. Yi Chuan Xue Bao 25:205–212

Part V
Case Studies

Phylogenetic Diversity of Culturable Marine Actinobacteria Isolated from the Havelock Island, the Andamans, India

Gobalakrishnan Rajagopal and Sivakumar Kannan

Abstract Due to the excellent track record of actinobacteria and their potential, more attention is being paid to the isolation of novel actinobacteria for the diverse range of screening programs. In this context, present study was aimed to investigate the phylogenetic diversity of culturable marine actinobacteria isolated from the coast of Havelock Island, the Andamans, India. Actinobacterial colonies were isolated from the sediments of the six stations of the Havelock Island, using different media, viz., Kuster's agar, Starch Casein agar, Actinomycetes isolation agar, Starch Casein nitrate agar, and AV agar. Higher actinobacterial population density was recorded in Kuster's agar. A total of 19 morphologically distinct actinobacterial strains were subjected to identification using cultural, cell wall chemotypical, morphological, physiological, and molecular (16S rDNA) characteristics; 18 out of 19 isolates belonged to *Streptomyces*, and 1 isolate belonged to *Actinoalloteichus*. This study has led us to conclude that the coastal sediments of the Havelock Island of the Andamans are good source for obtaining culturable marine actinobacteria. Isolation of such diverse marine actinobacteria from the marine sediments could yield novel candidates, a valuable source for various scientific and biotechnological applications and genes for their biosynthesis. Further research including molecular characterizations of the isolated strains are important to pursue their biotechnological importance and ecological roles.

Keywords Actinobacteria · Havelock Island · Cultivable diversity · *Streptomyces* · *Actinoalloteichus*

G. Rajagopal (✉)
Department of Biotechnology, J.J. College of Arts and Science (Autonomous), Pudukkottai, Tamilnadu, India

S. Kannan
Faculty of Marine Sciences, Centre of Advanced Study in Marine Biology, Annamalai University, Parangipettai, Tamilnadu, India

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_21

1 Introduction

The past 15 years have witnessed a dramatic increase in our knowledge of the crucial role played by the prokaryotes in early evolution (Doolittle 2000; Woese 2002; Martin and Embley 2004), and there is increasing recognition of the vast diversity and ecological importance of prokaryotes (Rappe and Giovannoni 2003). This is because of the better awareness of the role played by the biological communities in maintaining a sustainable biosphere as well as the rapid advances in molecular biology. An enormous amount of effort is being made worldwide by microbial ecologists to isolate and identify microorganisms from soil samples. In soil, estimates have revealed that 80–99% of the microorganisms remain unidentified (Leisack et al. 1997). Therefore, the challenge for the soil microbial ecologists is to identify the microbial populations and their guild that have key functional roles in specific soil processes.

Among the 18 major lineages of bacteria, phylum *Actinobacteria* is one of the largest units, and its divergence from other bacterial phyla is so ancient that it is currently not possible to identify their most closely related groups (Jami et al. 2015). Marine actinobacteria have been looked upon as potential sources of bioactive compounds, and the works done earlier have shown that these microbes are the richest sources of secondary metabolites. They hold the promising position as targets in screening programs due to their diversity and proven ability to produce novel metabolites (Manivasagan et al. 2013). However, their innovative bioactive characteristic has largely been overlooked in the past 150 years in human medicines (Seipke et al. 2012) with reference to antibacterial, antifungal, anticancer, antioxidant, and insecticidal activity in addition to enzyme inhibitions and production of novel enzymes (Sivakumar et al. 2007; Solanki et al. 2008). About 61% of the bioactive metabolites have been isolated from actinomycetes especially from streptomycetes and also from rare actinomycetes (non-streptomycetes) (Moncheva et al. 2002; Claverias et al. 2015). Hence, owing to their potential ability, actinobacteria are now given greater attention (Jami et al. 2015), and to identify new sources, many researchers focus on the new strains of this phylum (Zhang et al. 2016).

Actinobacteria are preliminarily identified according to their morphological criteria, including characteristics of colonies on the plate, morphology of substrate mycelium, aerial hyphae and spores, production of pigments, cell wall chemotype, whole-cell sugar pattern, and amino acids (Okami and Okazaki 1972; Hayakawa and Nonomura 1987a, 1987b; Nonomura 1988; Nolan and Cross 1988; Hayakawa et al. 1991; Cho et al. 1994; Kim et al. 1995; Sivakumar 2001). In the recent years, molecular phylogenetic analysis is being used to characterize microbial diversity, and investigators have turned to modern molecular tools based on PCR and phylogenetic analysis of the 16S rRNA gene for identification (Gray and Herwig 1996). Amplification of 16S rRNA genes using consensus microbial primers and separation of the resultant PCR amplicons by agarose gel electrophoresis can be done, and this information can be used to learn about the diversity of the habitat (Chowdhury et al. 2004). Modern taxonomy takes into account all available phenotypic and genotypic

data and integrates them in a consensus type of classification (Vandamme et al. 1996), and it aims at utilizing all the available data in delineating consensus group, decisive for the final conclusion (Gillis et al. 2001). Burkholder et al. (1954) had already emphasized that the microbial species especially actinomycetes should be characterized by multiple, readily recognizable, and reasonably stable properties. One characteristic alone, whatever it is, would be inadequate especially for the identification of a species (Krasilnikov 1960).

Islands represent ecologically sensitive and biogeographically significant ecosystems. These are the living laboratories on the earth for demonstrating the process of organic evolution (Das 2002). India has one of the major groups of islands, viz., the Andaman and Nicobar Islands, in the Bay of Bengal. The Andaman group of islands is adorned with several biotopes and is known for its rich marine wealth with mangroves, seagrasses, and coral reefs, which are the most productive ecosystems. Among the Andaman Islands, Havelock is a picturesque island with beautiful sandy beaches and lush green forests. It is a natural paradise, replete with white sand beaches, coral reefs with a colorful array of aquatic life, palm trees, as well as dense forests in the interiors (Gobalakrishnan et al. 2013). Exploration of microbial diversity in this island is very less. Hence, present study was made to explore the phylogenetic diversity of the cultivable marine actinobacteria from the coast of Havelock Island, the Andamans, India.

2 Material and Methods

2.1 Study Area (Fig. 1)

The Andaman group of islands in the Bay of Bengal lie between the latitude 11°97′ N and longitude 93°00′ E. In these island groups, Havelock Island, embracing mangrove, coral, seagrass, and beach habitats, was taken for our study. The sampling stations around this island were selected based on the different habitats: Station 1 (Mangrove 1, Lat. 12°02′27.9″ N and Long. 92°58′41.5″ E), Station 2 (EL Dorado—Coral 2, Lat 12°01′31.6″ N and Long 93°00′13.8″ E), Station 3 (Dive India—Beach 2, Lat 12°01′38.5″ N and Long 93°00′15.8″ E), Station 4 (Kalapathar—Mangrove 3, Lat 11°57′37.9″ N and Long 93°00′46.6″ E), Station 5 (Silver sand—Seagrass 1, Lat 12°00′36.1″ N and Long 93°00′31.0″ E), and Station 6 (Radhanagar Beach 3, 11°59′04.2″ N and 92°57′04.0″ E).

2.2 Collection of Samples

Field collections were made at six stations of the Havelock Island, during November 2011. Sediment samples were collected from here from a depth of 25 cm, using a sterile spatula. The samples were placed in sterile polythene covers and brought to

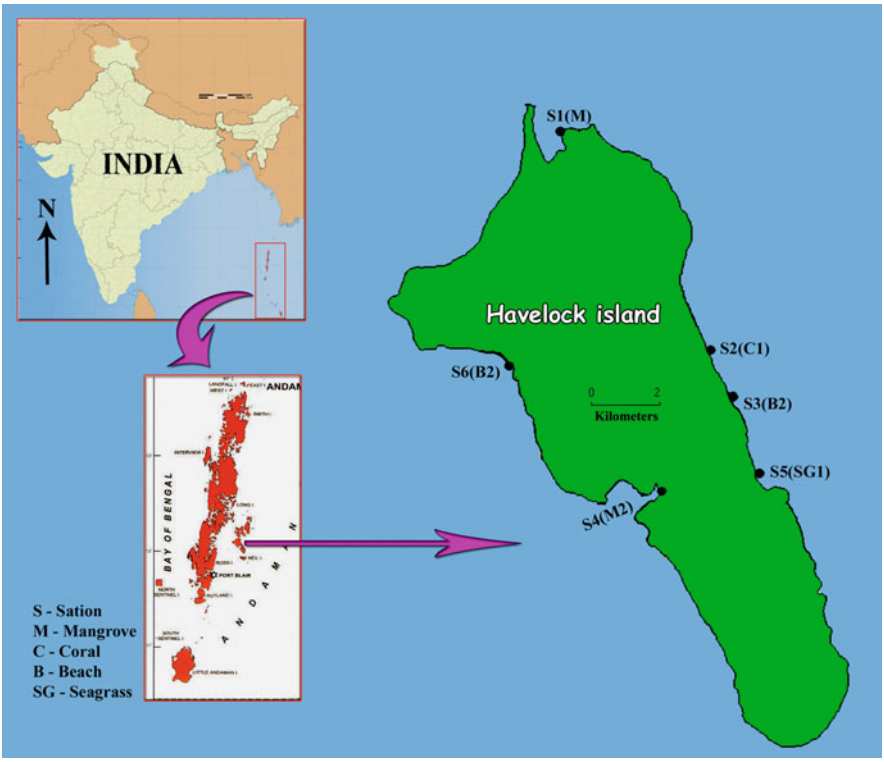


Fig. 1 Map showing the sampling stations in the Havelock Island

the field laboratory immediately, and, after arrival, necessary dilutions were made to carry out microbiological analysis.

2.3 Enumeration and Isolation of Actinobacteria

For the actinobacterial assessment, sediment samples were collected by inserting a polyvinyl corer (10 cm dia.) into the sediments. The corer was sterilized with alcohol before sampling at each station. The central portion of the top 2 cm sediment samples was taken out with the help of a sterile scoop. The sediment was then transferred to a sterilized polythene container and labeled. The sediment samples thus collected were air-dried aseptically. After a week, the sediment samples were mixed with equal weight of $CaCO_3$ (1:1) and incubated at 55 °C for 5 min (Balagurunathan 1992).

Actinobacterial population density was enumerated by adopting the spread plate method using different media, viz., Kuster's agar (KUA), Starch Casein agar (SCA), Actinomycetes isolation agar (AIA), Starch Casein nitrate agar (SCN), and AV agar (AVA); 1 g of pretreated sediment sample was serially diluted using sterile seawater, and 0.1 ml of serially diluted samples was added to the petri plates containing Kuster's agar medium and spread using a "L"-shaped glass spreader. The plates were incubated at 37 °C for 7 days in an inverted position. Leathery colonies of actinobacteria that appeared on the petri plates were counted from the 5th day onward up to 28th day. Population density of actinobacteria has been expressed as colony-forming units (CFU) per gram of the sample, in the text. All the colonies that grew on the petri plates were separately streaked and subcultured so as to ensure axenicity and were maintained in the slants.

2.4 Taxonomical Identification

2.4.1 Genus Level Identification

Characterization and subsequent identification of the strains to the genus level were made based on the chemical composition of the cell wall and micromorphological studies (Cummins and Harris 1958; Lechevalier and Lechevalier 1970; Tamura et al. 2000). Characterization of the potential strain MHA15 was made by following the method, described by Shirling and Gottlieb (1966), using the standard ISP medium.

Along with the cultural characteristics, melanoid pigments, reverse side pigments, soluble pigments, spore chain morphology, and assimilation of carbon sources were studied using the standard methods recommended by the International *Streptomyces* Project (Shirling and Gottlieb 1966).

2.4.2 Species-Level Identification

Species-level identification of the strain was made based on the keys and by using the *Bergey's Manual of Determinative Bacteriology*. The status of the strain was also confirmed based on its molecular characters (16S rDNA gene sequences), with the following methodology.

2.4.3 Genomic DNA Isolation and Amplification of 16S rDNA

Total genomic DNA was extracted from the actinobacterial broth by phenol-chloroform isoamyl alcohol method, which removes the protein and its cellular components from the nucleic acid to obtain the pure DNA (Ahmed et al. 2014). DNA sample (10 µl) was mixed with 2 µl of loading (6×) dye and loaded in 1% agarose gel. The separated DNA was visualized by UV transilluminator.

Each 50 µl amplification reaction contained 1 µl template DNA (50–200 ng), 5 µl 10× PCR buffer, 1 µl both forward [27F (5′-AGAGTTTGATCMTGGCTCAG-3′)] and reverse [1492R (5′-GGYTACCTTGTTACGACTT-3′)] primers, 1 µl dNTP mix (10 mM), 6 µl $MgCl_2$ (25 mM), 2.5 U Taq DNA polymerase, 2.5 µl DMSO, and 31.5 µl sterile water. The reaction conditions were initial denaturation at 95 °C for 30 s, annealing at 55 °C for 30 s and extension at 72 °C for 90 s. A final extension was performed at 72 °C for 10 min (Karuppiah et al. 2011). Reaction products were electrophoresed on a 1% agarose gel with 2 µl ethidium bromide and visualized under UV transilluminator and then purified.

2.4.4 16S rDNA Sequencing and Phylogenetic Analysis

The purified fragment was directly sequenced using a Ampli Tag FS DNA sequencing Kit (Applied Biosystem). The data were analyzed using applied Biosystem DNA editing, and assembly software and sequences comparisons were obtained using the MicroSEQ Software.

Sequence similarity search was made for the 16S rDNA sequence of the strain MHA15 by applying its sequence to BLAST search of the NCBI (National Center for Biotechnological Information, USA). Final editing of sequence alignment was done using BioEdit (Hall 1999), and phylogenetic analysis was performed with version 5 of the MEGA (Molecular Evolutionary Genetics Analysis) software package (Tamura et al. 2011). A phylogenetic tree was constructed, using the neighbor-joining tree-making algorithm (Saitou and Nei 1987). The topology of the phylogenetic tree was evaluated, using the bootstrap resampling method of Felsenstein (1985) with 1000 replicates.

3 Results

3.1 Actinobacterial Population Density

Actinobacterial population density ranged from 2×10^3 to 21×10^3 CFU g^{-1}. Higher actinobacterial population density (21×10^3 CFU g^{-1}) was recorded in Kuster's agar at station 4, and lower actinobacterial population density (2×10^3 CFU g^{-1}) was recorded in Actinomycetes isolation agar at station 5. In the Kuster's agar medium, higher actinobacterial population density (21×10^3 CFU g^{-1}) was recorded at station 4, and lower actinobacterial population density (2×10^3 CFU g^{-1}) was recorded at station 3. In the Starch Casein agar medium, higher actinobacterial population density (17×10^3 CFU g^{-1}) was recorded at station 4, and lower actinobacterial population density (3×10^3 CFU g^{-1}) was recorded at station 3. In the actinomycetes isolation agar medium, higher actinobacterial population density (9×10^3 CFU g^{-1}) was recorded at station 4, and lower actinobacterial population density (2×10^3 CFU g^{-1}) was recorded at station 5. In the Starch Casein nitrate agar medium, higher actinobacterial population

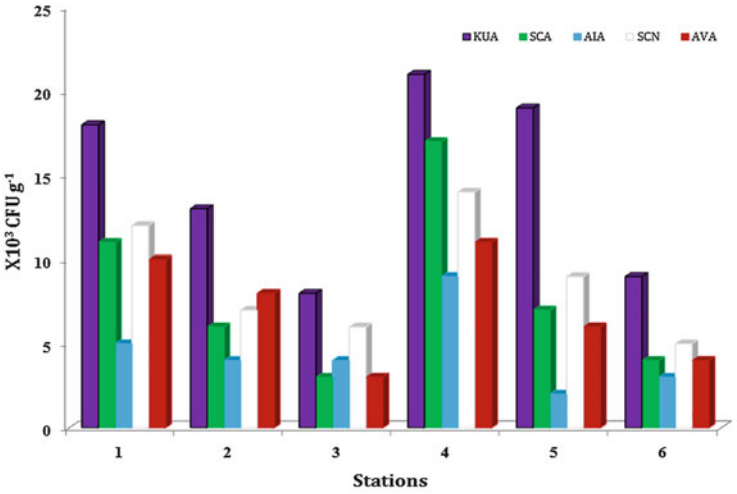


Fig. 2 Sediment actinobacterial population density recorded at six stations of the Havelock Island

density (14×10^3 CFU g^{-1}) was recorded at station 4, and lower actinobacterial population density (5×10^3 CFU g^{-1}) was recorded at station 6. In the AV agar medium, higher actinobacterial population density (11×10^3 CFU g^{-1}) was recorded at station 4, and lower actinobacterial population density (3×10^3 CFU g^{-1}) was recorded at station 3 (Fig. 2).

From the actinobacterial colonies isolated from the sediment samples of the six stations of the Havelock Island using various agar media, a total of 19 morphologically distinct actinobacterial strains were selected. These strains were isolated from stations 1 to 6, and the number of strains was 4, 2, 2, 5, 4, and 2, respectively, and all the 19 strains were labeled as MHA1, MHA2, MHA3, MHA4, MHA5, MHA6, MHA7, MHA8, MHA9, MHA10, MHA11, MHA12, MHA13, MHA14, MHA15, MHA16, MHA17, MHA18, and MHA19.

3.2 Identification of Actinobacteria

Phenotypic, chemotypic, and microscopic characters of the 19 isolates are given in Table 1, and physiological and biochemical characteristics are shown in Table 2. Out of the 19 morphologically distinct strains (MHA1 to MHA19), isolated in the present study, all the strains (18) except MHA15 showed the presence of LL-DAP along with glycine of the peptidoglycan layer with no characteristic sugar pattern, indicating that these strains belong to the cell wall chemotype I. In MHA15, cell wall characteristics showed the presence of glucose, mannose, galactose, and rhamnose as whole-cell sugars and meso-diaminopimelic acid as the amino acid in the cell wall; glycine was absent, indicating that this strain belongs to the cell wall chemotype III; 18 strains

Table 1 Phenotypic, chemotypical, and microscopic characteristics of isolated actinobacteria

Sl. No.	Isolate	Aerial mass color	Melanoid pigment	Reverse side pigment	Soluble pigment	Cell wall type	Spore chain nature	Genus
1	MHA1	White	–	–	–	I	Spirales	Streptomyces
2	MHA2	White	–	–	–	I	Spirales	Streptomyces
3	MHA3	White	–	+	–	I	Spirales	Streptomyces
4	MHA4	Gray	–	–	–	I	Rectiflexibiles	Streptomyces
5	MHA5	Whitish gray	–	+	–	I	Rectiflexibiles	Streptomyces
6	MHA6	Whitish yellow	–	–	–	I	Rectiflexibiles	Streptomyces
7	MHA7	Whitish yellow	–	+	+	I	Rectiflexibiles	Streptomyces
8	MHA8	Gray	–	+	–	I	Rectiflexibiles	Streptomyces
9	MHA9	Gray	–	–	–	I	Spirales	Streptomyces
10	MHA10	Whitish yellow	–	+	+	I	Rectiflexibiles	Streptomyces
11	MHA11	Whitish yellow	–	–	–	I	Spirales	Streptomyces
12	MHA12	Whitish yellow	–	–	–	I	Rectiflexibiles	Streptomyces
13	MHA13	White	–	+	–	I	Spirales	Streptomyces
14	MHA14	Gray	–	–	–	I	Spirales	Streptomyces
15	MHA15	Bluish gray	+	+	+	III	Long spore chain	Actinoalloteichus
16	MHA16	Whitish yellow	–	–	–	I	Spirales	Streptomyces
17	MHA17	White	–	+	–	I	Spirales	Streptomyces
18	MHA18	White	–	–	–	I	Spirales	Streptomyces
19	MHA19	Gray	–	–	–	I	Rectiflexibiles	Streptomyces

(+) Presence, (−) Absence

Table 2 Biochemical and physiological characteristics of isolated actinobacteria

Isolate	Gram's stain	Utilization as sole carbon source							
		Arabinose	Xylose	Inositol	Mannitol	Fructose	Rhamnose	Sucrose	Raffinose
MHA1	Gram + Ve	+	+	+	+	+	−	+	±
MHA2	Gram + Ve	−	±	+	−	+	−	+	±
MHA3	Gram + Ve	+	+	+	−	±	−	−	−
MHA4	Gram + Ve	+	+	+	+	+	+	+	−
MHA5	Gram + Ve	+	±	−	−	+	±	−	+
MHA6	Gram + Ve	−	±	+	+	+	−	−	±
MHA7	Gram + Ve	±	+	−	−	+	−	±	+
MHA8	Gram + Ve	+	+	±	−	±	−	+	+
MHA9	Gram + Ve	±	−	+	±	+	−	+	−
MHA10	Gram + Ve	−	+	−	−	+	−	±	+
MHA11	Gram + Ve	+	+	+	−	−	±	−	±
MHA12	Gram + Ve	+	+	−	+	+	−	±	−
MHA13	Gram + Ve	+	+	±	−	±	−	+	+
MHA14	Gram + Ve	±	+	−	+	±	+	−	+
MHA15	Gram + Ve	+	+	+	+	+	+	+	−
MHA16	Gram + Ve	±	±	−	+	+	−	+	−
MHA17	Gram + Ve	−	+	−	±	+	−	+	−
MHA18	Gram + Ve	−	+	+	+	+	+	−	−
MHA19	Gram + Ve	+	+	−	−	+	−	−	−

(continued)

Table 2 (continued)

Isolate	Gram's stain	Utilization as sole nitrogen source				
		Adenine	L-asparagine	L-cysteine	DL-methionine	L-phenylalanine
MHA1	Gram + Ve	+	±	±	+	−
MHA2	Gram + Ve	−	ND	−	ND	+
MHA3	Gram + Ve	−	+	±	−	−
MHA4	Gram + Ve	ND	+	−	+	±
MHA5	Gram + Ve	+	−	−	ND	+
MHA6	Gram + Ve	+	+	+	−	−
MHA7	Gram + Ve	±	+	ND	+	−
MHA8	Gram + Ve	±	+	−	+	+
MHA9	Gram + Ve	ND	−	+	+	ND
MHA10	Gram + Ve	ND	±	+	±	−
MHA11	Gram + Ve	+	−	±	+	+
MHA12	Gram + Ve	±	+	+	−	ND
MHA13	Gram + Ve	−	−	+	+	−
MHA14	Gram + Ve	−	+	−	−	ND
MHA15	Gram + Ve	−	−	+	−	+
MHA16	Gram + Ve	+	+	−	−	+
MHA17	Gram + Ve	±	±	ND	+	−
MHA18	Gram + Ve	+	+	+	±	ND
MHA19	Gram + Ve	ND	−	−	+	±

Positive (+); negative (−); weakly utilized (±); not detected (ND)

Table 3 16S rRNA gene taxonomic affiliation of cultivated strains obtained from the Havelock Island

Serial no.	Isolated strain name	Accession number	Sequence length (bp)	% sequence similarity	Closest type strain
1	MHA1	KF668652	1464	99.7	*S. chrestomyceticus*
2	MHA2	KF668653	0731	99.6	*S. sedi*
3	MHA3	KF668654	1342	100	*S. asiaticus*
4	MHA4	KF668655	1373	100	*Streptomyces* sp.
5	MHA5	KF668656	1460	100	*Streptomyces* sp.
6	MHA6	KF668657	1457	99.8	*S. griseobrunneus*
7	MHA7	KF668665	1373	100	*S. somaliensis*
8	MHA8	KF668658	0689	98.3	*S. finlayi*
9	MHA9	KF668666	1400	99.2	*S. albulus*
10	MHA10	KF681519	1438	98.4	*S. puniceus*
11	MHA11	KF668667	0926	97.9	*S. chungwhensis*
12	MHA12	KF668660	1398	100	*S. griseus*
13	MHA13	KF668661	1342	100	*S. cangkringensis*
14	MHA14	KF668662	1330	100	*S. hygroscopicus*
15	MHA15	KF668663	0873	91.6	*A. cyanogriseus*
16	MHA16	KF668668	1419	99.6	*S. cavourensis*
17	MHA17	KF668664	1330	99.9	*S. tsusimaensis*
18	MHA18	KF724893	1228	100	*S. albus*
19	MHA19	KF724894	0785	100	*Streptomyces* sp.

were identified as belonging to *Streptomyces* and 1 strain, MHA15, to *Actinoalloteichus*.

3.3 Phylogenetic Analysis of Actinobacteria

All the 19 strains were also subjected to molecular taxonomical analysis; their genomic DNA was extracted, purified, and subjected to amplification using thermo-cycler, and nearly full-length sequence of the 16S rRNA gene was obtained. Amplified products were sequenced, and all the sequenced data were deposited in NCBI GenBank, USA, and accession numbers were obtained. Based on the 16S rDNA sequences, phylogenetic tree was constructed with all the 19 strains (MHA1 to MHA19) and the previously reported sequences of *Streptomyces* species deposited in GenBank. Percentage of 16S rRNA gene sequence similarity of the isolates to their closest matches is presented in Table 3.

Phylogenetic tree revealed that the strains MHA1, MHA2, MHA3, MHA4, MHA5, MHA6, MHA7, MHA8, MHA9, MHA10, MHA11, MHA12, MHA13, MHA14, MHA16, MHA17, MHA18, and MHA19 closely resembled the previously

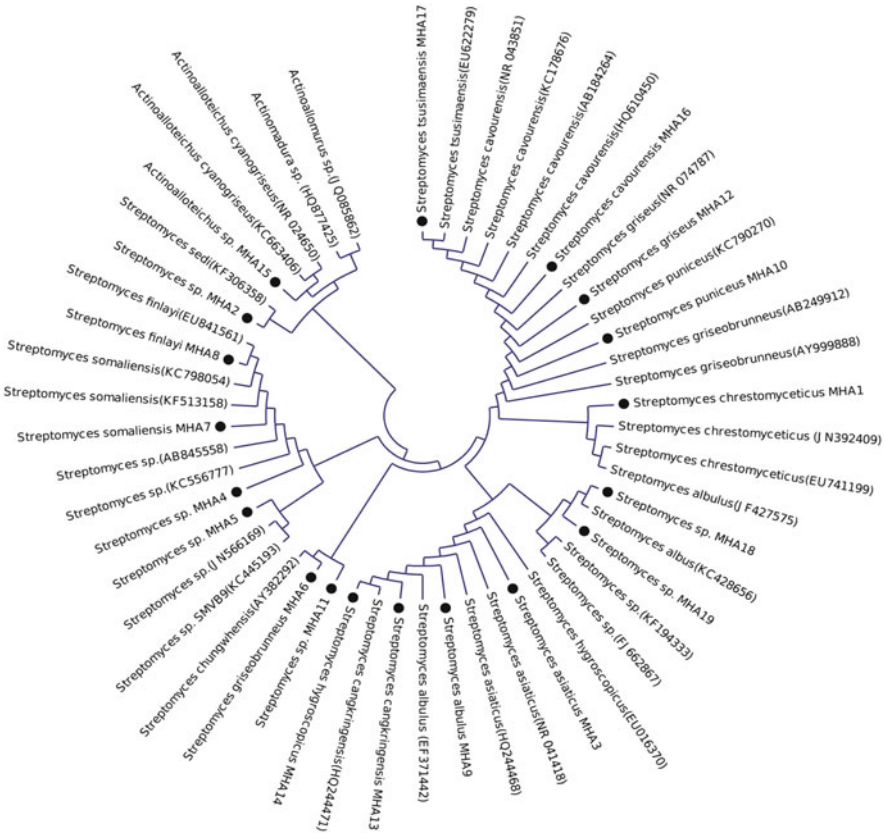


Fig. 3 Phylogenetic tree of actinobacterial isolates obtained from the coast of the Havelock Island. The tree was constructed based on the 16S rRNA gene sequencing neighbor-joining method with the percentage of bootstrap replicates (1000 resamplings)

reported *Streptomyces* species, and the strain MHA15 alone closely resembled the previously reported *Actinoalloteichus* species, as detailed above (Fig. 3).

MHA1, *Streptomyces chrestomyceticus*; MHA2, *S. sedi*; MHA3, *S. asiaticus*; MHA4, *Streptomyces* sp.; MHA5, *Streptomyces* sp.; MHA6, *S. griseobrunneus*; MHA7, *S. somaliensis*; MHA8, *S. finlayi*; MHA9, *S. albulus*; MHA10, *S. puniceus*; MHA11, *S. chungwhensis*; MHA12, *S. griseus*; MHA13, *S. cangkringensis*; MHA14, *S. hygroscopicus*; MHA16, *S. cavourensis*; MHA17, *S. tsusimaensis*; MHA18, *S. albus*; and MHA19, *Streptomyces* sp., and MHA15, *Actinoalloteichus* species.

4 Discussion

Recent past has witnessed a dramatic increase in our knowledge of the crucial role played by the prokaryotes in early evolution and a recognition of the vast diversity and ecological importance of prokaryotes (Gobalakrishnan 2015). This is due to a better understanding of the role of biological communities in maintaining a sustainable biosphere as well as the rapid advances in molecular biology. Efforts are being made worldwide by the microbial ecologists to isolate and identify soil microbes (Borneman et al. 1996). The challenge for the soil microbial ecologists is to identify the key functional roles of microbes in specific soil processes (Dastager et al. 2009). Among all the microbes, actinobacteria are widely distributed in the marine sediments as a small but significant fraction, with higher diversity (Ward and Bora 2006), and there is a lack of their taxonomic study worldwide (Oren and Stackebrandt 2002) to identify them. Claverias et al. (2015) have mentioned that marine sediments may harbor a great diversity of culturable actinobacteria. Therefore, present study was carried out to identify the cultivable marine actinobacteria through the modern taxonomical approach, including the culturable, chemotaxonomical, micromorphological, physiological, and molecular methods.

In the present study, 19 actinobacterial strains were isolated from the sediments of the Havelock Island using Kuster's agar as a selective medium, prepared in 50% of seawater, as this medium supports the isolation of various types of actinobacteria (Sivakumar 2001). Saadoun and Al-Momani (1996) found dominance of gray series *Streptomycetes* in Jordan soils, and the author isolated 105 gray-pigmented species of *Streptomyces*, using Kuster's agar medium. Similarly, Sahu et al. (2005) isolated higher number of actinobacteria using this medium from the Vellar estuary. Raghavendrudu and Kondalarao (2007) studied the distribution of actinobacteria in the Gaderu mangroves of Gautami-Godavari estuarine system, east coast of India. They used five different agar media for isolation, and, among them, Kuster's agar was more suitable for the isolation of the genus *Streptomyces.* Baskaran et al. (2011) also reported that Kuster's agar is a well supporting medium for marine actinobacteria. Sethubathi et al. (2013) have also reported that Kuster's agar medium has yielded higher counts of actinobacterial colonies and Mohseni et al. (2013) have isolated 44 actinobacterial strains from the sediments of the Caspian Sea, using Kuster's agar as one of the media. Very recently, Rajkumar et al. (2016) have reported higher actinobacterial population diversity (67×10^2 CFU g^{-1}) in Kuster's agar from the seagrass rhizosphere soil of the Gulf of Mannar Biosphere Reserve, India.

Chemical characters are being used increasingly in the classification and identification of microbes, and they have been responsible for some of the improvements in the taxonomy of actinomycetes (Lechevalier and Lechevalier 1970), and examination of whole-cell hydrolysates will give sufficient data for accurate identification. Many researchers have also found that the study of cell wall composition has good value in classification (Gerencser and Slack 1967; Kwapinski 1964; Pine and Boone 1967; Boone and Pine 1968). In addition, analysis of cell wall amino acids and whole-cell sugar would provide with the doorstep for the actinobacteriologists to

identify these organisms at the genus level (Rajkumar et al. 2016). Hence, presently isolated 19 actinobacterial strains were subjected to the analysis of whole-cell wall sugars and amino acids to identify them. Out of them, 18 strains showed the presence of LL-DAP along with glycine of the peptidoglycan layer with no characteristic sugar pattern, indicating that these strains belong to the cell wall chemotype I and the genus *Streptomyces*. Only one strain's cell wall characteristics showed presence of glucose, mannose, galactose, and rhamnose as whole-cell sugars and meso-diaminopimelic acid as the amino acid and glycine were absent, indicating that this strain belongs to the cell wall chemotype III and the genus *Actinoalloteichus*.

Broad actinomycetes groupings allow placement of isolates into reasonably well-defined categories based on color of the aerial mycelium (Pridham 1965). In addition, morphological features of the aerial mycelium are regarded as more significant for characterization, and they include the mode of branching and configuration of the spore chains (Anderson and Wellington 2001). In the present study, majority of the isolates produced aerial coiled mycelia, and the spores were arranged in spirales (11 isolates), rectiflexibles (7 isolates), and long chains (1 isolate). In the 19 isolates, color of the aerial mycelium was white (6 isolates), whitish yellow (6 isolates), gray (5 isolates), whitish gray (1 isolate), and blue gray (1 isolate).

Genus level identification of microorganisms by 16S rRNA gene sequencing has emerged as a more objective, accurate, and reliable method, with the added capability of defining taxonomical relationships among bacteria (Clarridge 2004), and a modern approach including physiological and chemotaxonomic analysis as well as phylogenetic tree analysis based on 16S rRNA gene sequences will lead to the identification of actinobacterial strains (Magarvey et al. 2004). Application of both traditional techniques, such as micromorphological, chemosystematic, and numerical taxonomy of phenotypic measurements and the modern molecular systematic techniques, is instrumental in clarifying the relationships within the actinomycetes (Stackebrandt et al. 1997; Atalan et al. 2000). In addition, current taxonomical approach replacing minimal numbers of characteristics by large number of features and characterizing different aspects of bacterial cells has resulted in more stability in the classification (Gillis et al. 2001). Further, it is widely recognized that multi-characteristic taxonomic approaches are better, since they take into account all available phenotypic and genotypic data and integrate them in a consensus type of classification (Vandamme et al. 1996). Present study was hence pursued with a combined taxonomical approach, viz., cultural, cell wall chemotypical, morphological, physiological, and molecular characteristics for the identification of the 19 isolates (MHA1 to MHA19), collected from the sediments of the Havelock Island, the Andamans.

Based on the above taxonomical approach, it was found that the stains MHA1, MHA2, MHA3, MHA4, MHA5, MHA6, MHA7, MHA8, MHA9, MHA10, MHA11, MHA12, MHA13, MHA14, MHA16, MHA17, MHA18, and MHA19 belong to the genus *Streptomyces*, and the strain MHA15 belongs to the genus *Actinoalloteichus*. For all the 19 isolates, individual sequences were submitted to the GenBank (National Center for Biotechnology Information, USA), and the accession

numbers were obtained. All the 19 strains were identified to their closest representative species.

So far, many species of actinobacteria belonging to 28 genera have been recorded from the marine environment. Among them, *Streptomyces* species are dominant (Karthik et al. 2010), as was found in the present study, where 18 out of 19 isolates belong to *Streptomyces* and 1 isolate belongs to *Actinoalloteichus*. In addition, traditionally *Streptomyces* species are considered as soil-dwelling organisms, and the success of these filamentous bacteria in the soil environment has been attributed to their ability to produce an array of catabolic enzymes that degrade biopolymers and also a mixture of antimicrobial compounds, depending on their substrates. Of the scores of soilborne microorganisms, *Streptomyces* species have been found to be the most prolific producers of a variety of clinically important biochemicals (Tarkka et al. 2008). In the present study, Havelock Island coastal sediments have recorded 95% of *Streptomyces* and 5% of *Actinoalloteichus*.

In conclusion, it can be stated that the coastal environment of the Havelock Island of the Andamans was found to harbor a good diversity of culturable marine actinobacteria with the representatives of *Streptomyces* and *Actinoalloteichus*. The 19 selected isolates were grouped into two phylotypes representing two families (*Streptomycetaceae* and *Pseudonocardiaceae*). Isolation of marine actinobacteria from the Havelock Island has indicated that novel and rare microbes, representing a valuable source for various scientific and biotechnological applications, could be obtained from marine sediments.

Acknowledgments The authors thank the Dean and Faculty of Marine Sciences and the authorities of Annamalai University, and one of the authors (R.G.) thank the Principal and the authorities of J.J. College of Arts and Science, for providing with necessary facilities. They also thank Prof. L. Kannan, Former Vice Chancellor, Thiruvalluvar University, Vellore, for encouragement and editorial support.

Conflict of Interest We declare that we have no conflict of interest.

References

Ahmed OB, Asghar AH, Elhassan MM (2014) Comparison of three DNA extraction methods for polymerase chain reaction (PCR) analysis of bacterial genomic DNA. Afr J Microbiol Res 8 (6):598–602

Anderson AS, Wellington MHE (2001) The taxonomy of *Streptomyces* and related genera. Int J Syst Evol Microbiol 51(3):797–814. https://doi.org/10.1099/00207713-51-3-797

Atalan E, Manfio GP, Ward AC, Kroppenstedt RM, Goodfellow M (2000) Biosystematic studies on novel *Streptomycetes* from soil. Antonie van Leeuwenhoek 77(4):337–353. https://doi.org/10.1023/A:1002682728517

Balagurunathan R (1992) Antagonistic actinomycetes from Indian shallow sea sediments with reference to α, β – unsaturated γ – lactone type of antibiotic from Streptomyces griseobrunneus. Ph.D. Thesis, Annamalai University, India

Baskaran R, Vijayakumar R, Mohan PM (2011) Enrichment method for the isolation of bioactive actinomycetes from mangrove sediments of Andaman islands, India. Mal J Microbiol 7 (1):26–32

Boone CJ, Pine L (1968) Rapid method for characterization of actinomycetes by cell wall composition. Appl Microbiol 16(2):279–284

Borneman J, Skroch PW, O'Sullivan KM, Palus JA, Rumjanek NG, Jansen JL, Nienhuis J, Triplett EW (1996) Molecular microbial diversity of an agricultural soil in Wisconsin. Appl Environ Microbiol 62(6):1935–1943

Burkholder PR, Sun SH, Ehrlich J, Anderson JL (1954) Criteria of speciation in the genus *Streptomyces*. Ann N Y Acad Sci 60(1):102–123. https://doi.org/10.1111/j.1749-6632.1954.tb40001.x

Cho SH, Hwang CW, Chung HK, Yang CS (1994) A new medium for the selective isolation of soil actinomycetes. J Appl Microbiol Biotechnol 22:561–563

Chowdhury SP, Khanna S, Verma SC, Tripathi AK (2004) Molecular diversity of tannic acid degrading bacteria isolated from tannery soil. J Appl Microbiol 97(6):1210–1219

Clarridge JE (2004) Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin Microbiol Rev 17(4):840–862

Claverias FP, Undabarrena A, González M, Seeger M, Cámara B (2015) Culturable diversity and antimicrobial activity of Actinobacteria from marine sediments in Valparaíso bay, Chile. Front Microbiol 6(737):1–11. https://doi.org/10.3389/fmicb.2015.00737

Cummins CS, Harris H (1958) Studies on the cell-wall composition and taxonomy of Actinomycetales and related groups. Microbiology 18(1):173–189

Das AK (2002) Islands. In: Alferd JRB, Das AK, Sanyal AK (eds) Ecosystem of India. ENVIS Center, Zoological Survey of India, Kolkata, pp 317–347

Dastager SG, Lee JC, Ju YJ, Park DJ, Kim CJ (2009) *Leifsonia kribbensis* sp. nov., isolated from soil. Int J Syst Evol Microbiol 59(pt1):18–21

Doolittle WF (2000) The nature of the universal ancestor and the evolution of the proteome. Curr Opin Struct Biol 10(3):355–358

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39(4):783–791

Gerencser MA, Slack JM (1967) Isolation and characterization of *Actinomyces propionicus*. J Bacteriol 94(1):109–115

Gillis M, Vandamme P, De Vos P, Swings J, Kersters K (2001) Polyphasic taxonomy. In: Boone DR, Castenholz RW (eds) Bergey's Manual® of systematic bacteriology, vol 1. Springer, New York, pp 43–48

Gobalakrishnan R (2015) Ecology, diversity and bioelectricity potential of marine actinobacteria from the Havelock island of the Andamans, India. Ph.D. Thesis, Annamalai University, India, p 230

Gobalakrishnan R, Kamala K, Poongodi S, Sivakumar K, Kannan L (2013) Microbial status of the coastal habitats in the Andamans, India. Int J Rec Scienti Res 4(5):506–514

Gray JP, Herwig RP (1996) Phylogenetic analysis of the bacterial communities in marine sediments. Appl Environ Microbiol 62(11):4049–4059

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95–98

Hayakawa M, Nonomura H (1987a) Humic acid-vitamin agar, a new medium for the selective isolation of soil actinomycetes. J Ferment Technol 65(5):501–509

Hayakawa M, Nonomura H (1987b) Efficacy of artificial humic acid as a selective nutrient in HV agar used for the isolation of soil actinomycetes. J Ferment Technol 65(6):609–616

Hayakawa M, Sadakata T, Kajiura T, Nonomura H (1991) New methods for the highly selective isolation of *Micromonospora* and *Microbispora* from soil. J Ferment Bioeng 72(5):320–326

Jami M, Ghanbari M, Kneifel W, Domig KJ (2015) Phylogenetic diversity and biological activity of culturable Actinobacteria isolated from freshwater fish gut microbiota. Microbiol Res 175:6–15

Karthik L, Kumar G, Bhaskara Rao KV (2010) Diversity of marine actinomycetes from Nicobar marine sediments and its antifungal activity. Int J Pharm Pharm Sci 2(1):199–203

Karuppiah V, Aarthi C, Sivakumar K (2011) Enhancement of PCR amplification of actinobacterial 16S rRNA gene using an adjuvant, dimethyl sulphoxide. Curr Sci 101(1):22–23

Kim CJ, Lee KH, Shimazu A, Kwon PDJ (1995) Isolation of rare actinomycetes from various types of soil. Kor J Appl Microbiol Biotechnol 23(1):36–42

Krasilnikov NA (1960) Rules for the classification of antibiotic producing actinomycetes. J Bacterial 76:75–80

Kwapinski JB (1964) Antigenic structure of the *actinomycetales* VII. Chemical and serological similarities of cell walls from 100 *actinomycetales* strains. J Bacteriol 88(5):1211–1219

Lechevalier MP, Lechevalier H (1970) Chemical composition as a criterion in the classification of aerobic actinomycetes. Int J Syst Evol Bacteriol 20(4):435–443. https://doi.org/10.1099/00207713-20-4-435

Leisack W, Jessen PH, Rainey FA, Ward-Rainey N, Stackebrandt E (1997) Microbial diversity in soil: the need for a combined approach using molecular and cultivation techniques. In: Elsas JDV, Trevors JT, Wellington EMH (eds) Modern soil microbiology. Wellington Marcel Dekker Inc., New York, pp 375–439

Magarvey NA, Keller JM, Bernan V, Dworkin M, Sherman DH (2004) Isolation and characterization of novel marine-derived actinomycete taxa rich in bioactive metabolites. Appl Environ Microbiol 70(12):7520–7529

Manivasagan P, Venkatesan J, Sivakumar K, Kim SK (2013) Marine actinobacterial metabolites: current status and future perspectives. Microbiol Res 168(6):311–332

Martin W, Embley TM (2004) Evolutionary biology: early evolution comes full circle. Nature 431:134–137

Mohseni M, Norouzi H, Hamedi J, Roohi A (2013) Screening of antibacterial producing actinomycetes from sediments of the Caspian Sea. IJMCM 2(6):64–71

Moncheva P, Tishkov S, Dimitrova N, Chipeva V, Nikolova SA, Bogatzevska N (2002) Characteristics of soil actinomycetes from Antarctica. J Cult Coll 3:3–14

Nolan RD, Cross T (1988) Isolation and screening of actinomycetes. In: Goodfellow M, Willams ST, Mordarski M (eds) Actinomycetes in biotechnology. Academic Press, London, pp 1–32

Nonomura H (1988) Isolation, taxonomy and ecology of soil actinomycetes. Actinomycetologica 3:45–54

Okami Y, Okazaki T (1972) Studies of marine microorganisms. II. Actinomycetes in Sagami bay and their antibiotic substance. J Antibiot 25(8):456–460

Oren A, Stackebrandt E (2002) Prokaryote taxonomy online: challenges ahead. Nature 419 (6902):15. https://doi.org/10.1038/419015c

Pine L, Boone CJ (1967) Comparative cell wall analysis of morphological forms within the genus *Actinomyces*. J Bacteriol 94(4):875–883

Pridham TG (1965) Color and Streptomycetes. Report of an international workshop on determination of color of Streptomycetes. Appl Microbiol 13:43–61

Raghavendrudu G, Kondalarao B (2007) Studies on the distribution of actinobacteria in the Gaderu mangroves of Gautami Godavari estuarine system, East coast of India. J Mar Biol Assoc India 49(2):246–249

Rajkumar J, Sivakumar K, Thangaradj T (2016) Characterization of Actinobacterial population in the Seagrasses Rhizosphere Soils of the Gulf of Mannar Biosphere Reserve, India. Res J Microbiol 11(1):1–10. https://doi.org/10.3923/jm.2016.1.10

Rappe MS, Giovannoni SJ (2003) The uncultured microbial majority. Annu Rev Microbiol 57:369–394

Saadoun I, Al-Momani F (1996) Bacterial and *Streptomyces* flora of some Jordan valley soils. Actinomycetes 7(3):95–99

Sahu MK, Sivakumar K, Kannan L (2005) Degradation of organic matters by the extra-cellular enzymes of actinomycetes isolated from the sediments and molluscs of the Vellar estuary. J Aqua Biol 20(2):142–144

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4):406–425

Seipke RF, Kaltenpoth M, Hutchings MI (2012) *Streptomyces* as symbionts: an emerging and widespread theme? FEMS Microbiol Rev 36:862–876. https://doi.org/10.1111/j.1574-6976.2011.00313.x

Sethubathi GVB, Sivakumar K, Thangaradjou T, Kannan L (2013) Ecology and population density of the marine actinobacteria of little Andaman island, India. Ind J Mar Sci 42(3):390–401

Shirling EB, Gottlieb D (1966) Methods for characterization of *Streptomyces* species. Int J Syst Evol Bacteriol 16(3):313–340. https://doi.org/10.1099/00207713-16-3-313

Sivakumar K (2001) Actinomycetes of an Indian mangrove (Pichavaram) environment: an inventory. Ph.D. Thesis, Annamalai University, India, p 91

Sivakumar K, Sahu MK, Thangaradjou T, Kannan L (2007) Research on marine actinobacteria in India. Ind J Microbiol 47(3):186–196

Solanki R, Khanna M, Lal R (2008) Bioactive compounds from marine actinomycetes. Ind J Microbiol 48(4):410–431

Stackebrandt E, Rainey FA, Ward-Rainey NL (1997) Proposal for a new hierarchic classification system, Actinobacteria classis nov. Int J Syst Bacteriol 47(2):479–491

Tamura T, Zhiheng L, Yamei Z, Hatano K (2000) *Actinoalloteichus cyanogriseus* gen. nov., sp. nov. Int J Syst Evol Microbiol 50(Pt4):1435–1440. https://doi.org/10.1099/00207713-50-4-1435

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10):2731–2739

Tarkka MT, Lehr NA, Hampp R, Schrey SD (2008) Plant behavior upon contact with *Streptomycetes*. Plant Signal Behav 3(11):917–919

Vandamme P, Pot B, Gillis M, De Vos P, Kersters K, Swings J (1996) Polyphasic taxonomy, a consensus approach to bacterial systematics. Microbiol Rev 60:407–438

Ward AC, Bora N (2006) Diversity and biogeography of marine actinobacteria. Curr Opin Microbiol 9(3):279–286

Woese CR (2002) On the evolution of cells. Proc Natl Acad Sci U S A 99(13):8742–8747

Zhang B, Xiukun W, Zhang G, Li S, Zhang W, Chen X, Sun L, Baogui Z, Guangxiu L, Chen T (2016) The diversity and biogeography of the communities of Actinobacteria in the forelands of glaciers at a continental scale. Environ Res Lett 11(5):1–12

DNA Barcoding and Molecular Phylogeny of Indigenous Bacteria in Fishes from a Tropical Tidal River in Malaysia

Mohammad Mustafizur Rahman, Mohd Haikal Izzuddin, Najmus Sakib Khan, Akbar John, and Mohd Azrul Naim

Abstract DNA barcoding along with molecular phylogeny can be used to taxonomic identification, characterization, and discovery of species and understand molecular relationships especially in term of species divergence. Thus, they facilitate biodiversity studies. Some studies have addressed DNA barcoding of bacterial samples from various sources. Unfortunately, the DNA barcoding of fish bacterial diversity has yet been studied especially in the tropical tidal river. Therefore, a study was conducted to (1) identify the observed bacterial isolates by comparing the partial sequence from an unknown sample to a collection of sequences from known reference samples, (2) know the taxonomic and phylogenetic identity of identified bacteria in the fish, and (3) know the abundance of fish bacteria in the Kuantan River. For this study, three commercially important fish, namely, *Pristipomoides filamentosus*, *Cyclocheilichthys apogon*, and *Labiobarbus festivus* were captured with gill nets from Kuantan River, Malaysia. Bacteria from skin, gill, and gut in fish were cultured at 35 °C for 24 h in both nutrient and marine agars. Bacterial DNA were extracted using Bacterial Genomic DNA Isolation Kit following manufacturer's specifications. Isolated DNA were quantified in NanoDrop 2000v and gel eluded in 1.5% agarose gel and visualized under gel visualizer. PCR products were out source sequenced at First Base Sdn.bhd using ABI sequencer under Sanger sequencing method. Sequences were trimmed using Sequence Scanner 2.0V. The aligned sequences were inspected by eye and edited to remove ambiguities based on PHRED scores and chromatogram. Fully aligned sequences were subjected to BLAST for nucleotide similarity search against 16S rRNA database. Best matched species was selected based on BLAST

M. M. Rahman (✉)
Department of Marine Science, Faculty (Kulliyyah) of Science (KOS), International Islamic University Malaysia (IIUM), Kuantan, Pahang, Malaysia

INOCEM Research Station, KOS, IIUM, Kuantan, Malaysia
e-mail: mustafiz@iium.edu.my

M. H. Izzuddin · N. S. Khan · M. A. Naim
Department of Biotechnology, KOS, IIUM, Kuantan, Pahang, Malaysia

A. John
INOCEM Research Station, KOS, IIUM, Kuantan, Malaysia

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_22

results and lowest genetic distance between known and unknown nucleotides. Genetic distance (sequence divergences) were calculated using the K2P (Kimura two parameter) distance model. Neighbor-joining (NJ) trees of K2P distances were created to provide a graphic representation of the patterning of divergence between species. This method identified a total of 11 fish bacteria, which taxonomically classified into Enterobacteriales, Pseudomonadales, Actinomycetales, and Bacillales. The range of pair-wise genetic distances between species of Enterobacteriales was lower than Bacillales. Similarly, within group mean genetic distance of Enterobacteriales (0.010) was lower than Bacillales (0.055). These results indicate that the identified bacteria species under Enterobacteriales are more closely related than bacteria species under Bacillales. The mean genetic distance between groups were genetically almost equally closed which was confirmed by overall mean diversity. Out of 11 species 7 were identified in *Cyclocheilichthys apogon*, 8 in *Labiobarbus festivus*, and 7 in *Pristipomoides filamentosus*. Overall mean bacterial abundance (CFU/g) were higher in *C. apogon* (6.68×10^3) compare to *L. festivus* (5.12×10^3) and *P. filamentosus* (5.20×10^3). Overall, highest bacterial abundances were observed in fish gut (6.62×10^3), followed by fish gill (5.78×10^3) and fish skin (4.60×10^3).

Keywords *Cyclocheilichthys apogon* · *Labiobarbus festivus* · *Pristipomoides filamentosus* · Bacteria · Kuantan River · Malaysia

1 Introduction

Bacteria are smallest and most active biological entities in all in aquatic environments. They play the most important role in mineralization of organic matter, and, thus, they execute biogeochemical cycles (Rahman et al. 2008a, b). Subsequently, primary, secondary, and tertiary production and aquatic food webs are executed (Rahman 2015a, b). Therefore, heterotrophic bacteria are a major constituent in all aquatic environments. Fish are continuously exposed to the bacteria present in the water, which influences the bacterial flora on external surfaces of fish (El-Shafai et al. 2004). Similarly, the digestive tract receives water and food that is populated with different types of bacteria. However, colonization of bacteria starts at the egg and/or larval stage and continue with the development of the fish (Olafsen 2001). Thus, the numbers and range of bacteria present in the eggs, on food, and in the water influence the bacterial flora of the developing fish. However, determining bacterial community composition in fish is one of the necessary steps in understanding fish bacterial ecology. Moreover, understanding abundance of pathogenic bacteria in fish may help to prepare a proper management policy to protect fish and human from diseases caused by bacteria.

Bacterial floras of fish are grossly two categories. Numerous bacteria cause diseases in fish and shellfish. In general, marine fishes are susceptible to diseases caused by *Vibrio anguillarum* and freshwater species to those caused by *Aeromonas hydrophila* and *A. salmonicida*. Majority of bacterial pathogens in fish are also

capable of causing diseases in humans. They produce toxins that cause many lethal diseases such as paralytic shellfish poisoning, neurotoxic shellfish poisoning, diarrheic shellfish poisoning, amnesiac shellfish poisoning, and ciguatera fish poisoning (Bienfang et al. 2011). Some bacteria play an important role with respect to the well-being and health of fish (Hasan et al. 2007). They help digestion process and control harmful microflora of fish (Robertson et al. 2000). For example, the bacteria *Vibrio alginolyticus* reduces diseases in Atlantic salmon (*Salmo salar*) caused by infection of common pathogenic bacteria (*Aeromonas salmonicida*, *Vibrio anguillarum*) (Austin et al. 1995). However, interest is increasingly growing in the society to know type and abundance of bacteria live in fish as they are important part of the human diet fish as it is the major sources of animal protein in many countries (Smith and Prairie 2004; Van Horn et al. 2016). It is one of the preferred human foods in many countries in the world. Besides high nutritional value, fish and fish products play a very important role against many diseases such as coronary heart disease, cancer, diabetes, and inflammatory disease (Rahman and Verdegem 2007; Rahman and Meyer 2009; Rahman et al. 2009).

Type and abundance of bacteria in fish are dependent on many factors such as species, feeding pattern, age, size, season, geographical location, and environmental condition (Novotny et al. 2004). There are some published information on the type and bacteria in river fishes. However, study on bacteria in fish from a tropical tidal river is limited as identification of bacteria is a very important issue in bacteriological study. Most of the studies identified river fish bacteria through standard morphological, physiological, and biochemical tests. For example, de Sousa and Silva-Souza (2001) observed bacterial community associated with Congonhas River fish morphological, physiological, and biochemical tests. However, morphological, physiological, and biochemical tests are unable to identify some groups of bacteria. According to Figueras et al. (2011), morphological and physiological characterizations can give imprecise results; one cannot guarantee that a bacteria strain has been correctly identified at the species level, if it has not been verified using a reliable molecular method. Moreover, these methods do not give any information about genetic closeness or divergence among various species of bacteria. This problem can be overcome by DNA barcoding method, which identifies bacteria using partial sequence of DNA (Hebert and Gregory 2005).

DNA barcoding can be used to diagnose taxa, increasing the speed, objectivity, and efficiency of species identification (Hickerson et al. 2006; Stahlhut et al. 2012). DNA barcoding and molecular phylogenetic analysis also helps to understand molecular relationships especially in term of species divergence among various fish bacteria (Hebert and Gregory 2005). According to Bhattacharya et al. (2016), DNA barcoding appears to be a promising approach for taxonomic identification, characterization, and discovery of newer diverged species and thus facilitating biodiversity studies. It helps researchers to appreciate genetic and evolutionary associations by collection of molecular, morphological, and distributional data. According to Begerow et al. (2010), DNA barcoding provides very important information to evaluate ecological sequences for resolving management priorities. Therefore, molecular phylogeny is necessary not only to identify species but also to

understand the genetic closeness or relationship among various bacteria species. In this study, fish bacteria were identified by DNA barcoding 16S rRNA sequencing as the method employed one or few relatively short gene sequence in the genome which is unique enough to identify species (Purty and Chatterjee 2016).

All fishes were collected from a tropical tidal river. Tide and salinity are two important factors which greatly influence the dynamics of fish bacteria in the tidal river. In this study, to examine the type and abundance of fish bacteria in a tropical tidal river, the Kuantan River was taken as a model as this river plays a life-sustaining role for many inhabitants in Kuantan and is important in economy, ecology, and recreation in the east coast of Peninsular Malaysia. The main objectives of the study were to (1) identify the observed bacterial isolates by comparing the partial sequence from an unknown sample to a collection of sequences from known reference samples, (2) know the taxonomic and phylogenetic identity of identified bacteria in the fish, and (3) know the abundance of fish bacteria in the Kuantan River.

2 Materials and Methods

2.1 Collection and Preparation of Samples

For bacteriological analysis, three commercially important fish, namely, *Pristipomoides filamentosus*, *Cyclocheilichthys apogon*, and *Labiobarbus festivus* were selected. *Pristipomoides filamentosus* is a saline water fish which was captured from Kuantan River estuary. *Cyclocheilichthys apogon* and *Labiobarbus festivus* are freshwater fishes and were collected from 15 km upstream of Kuantan River estuary where water salinity fluctuated between 0 and 5 °C. All fishes were captured with gill nets and immediately transported to the laboratory where they were dissected using scalpel, forceps, scissors, and knives. Skin, gills, and gut samples of fish were aseptically obtained by dissection. 1 g of each dissected organs were separately homogenized by a mortar and pestle. All apparatuses were sterilized prior to dissection.

2.2 Culture and Quantitative Estimation of Bacteria

The homogenized tissues from skin, gill, and gut were serially diluted and cultured at 35 °C for 24 h in both nutrient and marine agars. *Aseptic technique was incorporated in all steps of the bacteriological works.* To avoid problem regarding overcrowding of bacterial colonies on the agar plates, a few trials were conducted to find the best dilution method. All colonies were counted and expressed in CFU (colony-forming unit) per g of sample. All bacterial colonies isolated from skin, gill, and gut tissues were morphologically grouped on their colony shape, size, and color.

2.3 Isolation and Sequencing of DNA

A total of 14 morphologically different colonies (C1–C14) were identified. Each bacterial colony was subcultured and subjected to DNA isolation using Bacterial Genomic DNA Isolation Kit (Cat. No. 17900, Norgen Biotek Corp., Canada) following manufacturer's specifications. Isolated DNA were quantified in NanoDrop 2000v and gel eluded in 1.5% agarose gel and visualized under gel visualizer. Samples with low DNA concentration were further isolated to achieve ample DNA concentration for downstream applications. It can be noted that DNA isolation for C11 and C14 was unsuccessful even after multiple DNA isolation steps and hence excluded from the analysis. The extracted DNA was stored at −20 °C for further use. Isolated DNA was subjected to PCR under standard thermal cycle condition. PCR reaction mixture consisted of 1.5 mM $MgCl_2$, 0.2 mM of each DNTPs, 1U of Taq DNA polymerase, 0.2 μM of forward primer (27b F—5′ AGAGTTTGATCCTGGCTCAG 3′), and 2 μl of DNA template in a total volume of 25 μl. Following thermal cycle condition was used in amplifying target gene: 95 °C for 5 min for initial hot start, followed by 40 cycles at 95 °C for 30 s, 56 °C for 30 s for annihilation, and 72 °C for 1 min and a final extension at 72 °C for 10 min. The PCR products were gel eluded in 1.5% agarose gel for further documentation. PCR products were out source sequenced at First Base Sdn. Bhd. using ABI sequencer using Sanger sequencing method. Sequences were trimmed using Sequence Scanner 2.0V. Sequence for sample (C2) was not aligned as there were many mismatches and ambiguous base pairs observed throughout the sequence.

2.4 Data Analysis

Quantitative bacterial data were checked for normality using *Shapiro–Wilk test* and homogeneity of variance using *Levene's test*. They were subjected to one-way ANOVA (analysis of variance) to understand their variation in skin, gills, and gut of fishes. If an effect was significant, ANOVA was followed by Tukey's test for unplanned multiple comparisons of means. Statistically significance was determined at $P \leq 0.05$. All statistical analyses were performed using *IBM SPSS* software for Windows (version 22.0).

To understand the nucleotide similarity, all sequences were aligned using Sequence Scanner 2 software. The aligned sequences were inspected by eye and edited to remove ambiguities based on PHRED scores and chromatogram. Fully aligned sequences were subjected to BLAST for nucleotide similarity search against 16S rRNA database. For each sequence, a total of five closest nucleotides were selected to understand the closest species by genetic distances and phylogenetic tree using Mega 6 software (Tamura et al. 2013). Best matched species was selected based on BLAST results and lowest genetic distance between known and unknown nucleotides (Table 1). Genetic distance (sequence divergences) were calculated

Table 1 Best matched species and second best matched species based on 16S rRNA gene and its acquired GenBank accession numbers

Sample number	Best matched species				Second best matched species			
	Species	Accession number	I (%)	Source	Species	Accession number	I (%)	Source
C1	*Enterobacter aerogenes*	AM184247	99	River water	*Pantoea agglomerans*	KX684195	99	Wastewater
C3	*Leucobacter chromiireducens*	KT262983	97	River water	*Leucobacter salsicius*	NR117298	97	Fermented food
C4	*Edwardsiella tarda*	DQ855293	99	–	*Escherichia* sp.	KR063559	99	Wetland
C5	*Staphylococcus saprophyticus*	KT986100	100	Ocean water	*Staphylococcus saprophyticus*	KX343982	100	Acidic soil
C6	*Psychrobacter* sp.	AM422129	94	Drainage sample	*Psychrobacter faecalis*	KU364044	93	Seawater
C7	*Enterobacter cloacae*	KT328453	99	Intestine of fish	*Leclercia adecarboxylata*	NR104933	99	Drinking water
C8	*Enterobacter mori*	KF747680	99	Rhizosphere soil	*Enterobacter cloacae*	HM854373	99	Rhizosphere
C9	*Macrococcus caseolyticus*	KC969082	99	East China Sea mud	*Macrococcus caseolyticus*	KJ958217	99	Yellow croaker
C10	*Bacillus tequilensis*	KU356174	99	South China Sea	*Bacillus subtilis*	KT986107	99	Soil
C12	*Bacillus licheniformis*	KJ842640	99	Hot spring	*Bacillus licheniformis*	KT748523	99	Fecal matter
C13	*Bacillus pumilus*	KT986126	99	Ocean water	*Bacillus stratosphericus*	KM277362	99	Fish gut

I identity

using the K2P (Kimura two parameter) distance model (Kimura 1980). Neighbor-joining (NJ) trees of K2P distances were created to provide a graphic representation of the patterning of divergence between species. Before calculating genetic distances and constructing phylogenetic tree, all sequenced DNA were subjected to complete alignment using ClustalX software (version 2.1). By phylogenetic relatedness among known and unknown nucleotides based on the homology of 16S rRNA sequences, the closest affiliation of a new isolate was assigned. Phylogenetic relationship was observed within all known nucleotides to assign each species in a group. Neighbor-joining analysis was employed to calculate numerical pair-wise genetic distance.

3 Results and Discussion

This study provides qualitative and quantitative bacteriological information about fish bacteria in the Kuantan River. Qualitative and quantitative determination of bacterial communities are usually very challenging as only a small percentage of bacteria can be cultured under laboratory conditions. In this study, two types of agar media (nutrient and marine agars) were used to culture bacteria. Most important challenge of bacteriological study is taxonomical identification up to species level. Morphological, physiological, and biochemical tests are traditionally used to identify bacteria, but they are not precise methods to identify some groups of bacteria up to species level. To overcome this challenge, the DNA barcoding technique using genetic distance and phylogenetic tree was applied to identify bacteria up to species level. In the present study, to identify bacteria, a conserved region of 16S rRNA was amplified as these genes (1) are ubiquitous (present in all prokaryotes), (2) are structurally and functionally conserved, and (3) contain variable and highly conserved regions (Hugenholtz 2002). According to Alperi et al. (2008), the 16S rRNA gene is essential to identify bacteria as this gene can be more discriminative especially in closely related strains. Therefore, this gene is commonly used as a molecular marker to identify bacteria (Han et al. 2011; Hidalgo and Figueras 2012; Table 2).

Figure 1 shows the phylogenetic tree that indicates genetic relationship and bootstrap values among known and unknown DNA sequences. Species ecology along with lowest genetic distance and highest bootstrap value between known and unknown species led to identify a total 11 species of fish bacteria. The mean nucleotide composition of identified 11 species was estimated as T = 19.4, C = 22.9, G = 31.7, and A = 26.0%. The phylogenetic relationships and pair-wise genetic distance within identified bacteria are presented in Fig. 2 and Table 3, respectively. Phylogenetic tree divided all bacteria into four clades. Taxonomically, Clade 1, Clade 2, Clade 3, and Clade 4 belong to Enterobacteriales (*Enterobacter aerogenes*, *Enterobacter mori*, *Enterobacter cloacae*, and *Edwardsiella tarda*), Pseudomonadales (*Psychrobacter* sp.), Actinomycetales (*Leucobacter chromiireducens*), and Bacillales (*Bacillus tequilensis*, *Bacillus licheniformis*, *Bacillus pumilus*, *Macrococcus caseolyticus*, and *Staphylococcus saprophyticus*),

Table 2 Published information of DNA barcoding to identify various species of bacteria

Identifying bacteria	Gene name	Reference
Bacteria within *Enterobacteriaceae*	16S rRNA	Hauben et al. (1998)
Enterobacteriaceae (some genera)	16s rDNA	Sprber et al. (1998)
Cronobacter spp.	16S rRNA	Schmid et al. (2009)
Pantoea sp.	16S rRNA	Brady et al. (2008)
Pantoea agglomerans strains	16S rRNA	Deleetoile et al. (2009)
Erwinia and *Pantoea* species	Whole genome	Zhang and Qiu (2015)
Lactobacillus species	23S rRNA, 16S rRNA	Kwon et al. (2004)
Genera *Pectobacterium* and *Dickeya*	16S rRNA	Ma et al. (2007)
Enterobacteriaceae (some species)	16S rDNA	Drancourt et al. (2001)
Raoultella terrigena	16S rRNA	Wang et al. (2016)
Fish bacteria	16S rRNA	Alikunhi et al. (2017)
Fish spoilage bacteria	16S rDNA	Garcıa-Lopez et al. (2004)
River water bacteria	16S rRNA	Cottrell et al. (2005)
Psychrophilic bacteria	16S rDNA	Maruyama et al. (2000)
Fish bacteria	16S rRNA	Sebastiao et al. (2015)
Aeromonas (from drinking water)	16S rRNA	Alzahrani (2015)

respectively. The range of pair-wise genetic distances between species of Enterobacteriales (range, 0.001–0.015) was lower than Bacillales (range, 0.016–0.780) (Table 3). Similarly, within group mean genetic distance of Enterobacteriales (0.010) was lower than Bacillales (0.055). These results indicate that the identified bacteria species under Enterobacteriales were more closely related than bacteria species under Bacillales (Dunbar et al. 2002). The mean genetic distance between groups Enterobacteriales and Pseudomonadales was the lowest (0.229) compared to all other pair groups although the range mean genetic distance between groups was 0.029 and 0.331 (Table 4). This indicates that all groups were genetically almost equally closed and identified bacteria were genetically less diversified (Dunbar et al. 2002; Rastogi and Sani 2011). This can be further conformed by overall mean diversity among all identified bacterial species. We observed the mean species diversity among all identified bacterial species as 0.209. According to Rastogi and Sani (2011), the typical clone libraries of 16S rRNA genes contain lower number of sequences ($\leq$1000 sequences) and therefore reveal only a small portion of the microbial diversity present in a sample. Some studies have addressed DNA barcoding of bacterial samples from various sources (Table 2). Unfortunately, the DNA barcoding of fish bacterial diversity has yet been studied especially in the tropical tidal river. Studies on bacterial identification is usually employing either one of the two gene sequencing such as 16S rRNA or chaperonin protein (cpn60). However, due to larger barcode gap in cpn60, 16s rRNA gene is usually preferred by framework established by the International Barcode of Life (IBL) (Links et al. 2012). The method of barcoding 16s rRNA is reliable, accurate, and used to track novel microbes and further metagenomic analysis. Until recently, the link between phylogenetic hypotheses and questions raised by

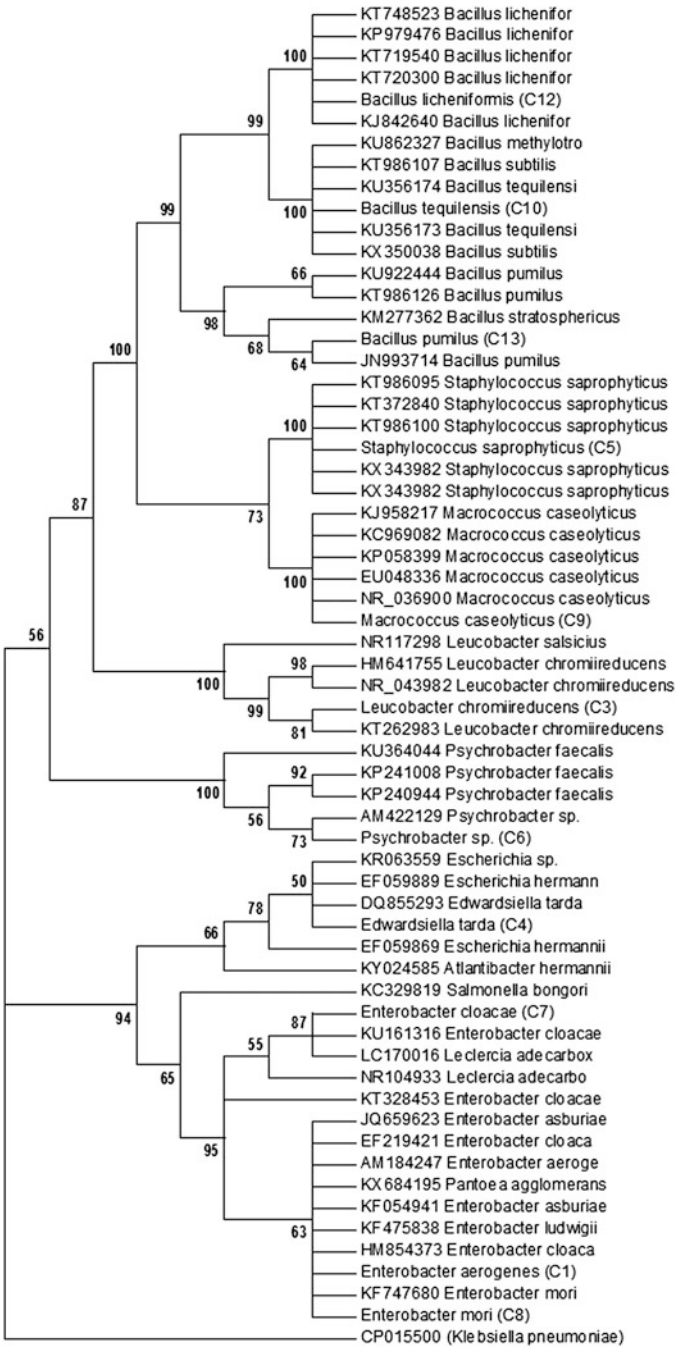


Fig. 1 Neighbor-joining dendrogram depicting the estimated phylogenetic relationships among unknown (indicating C1, C2, etc. in parentheses) and related known species (indicating with accession number) based on 16S rRNA GenBank. The optimal tree with the sum of branch length = 1.68418054 is shown. The percentage of replicate trees in which the associated taxa clustered

fundamental ecologists pertaining to the distribution and assemblage of endosymbiotic bacteria in host species is not answered. Recent studies have proven the coevolution of host and endosymbionts gene at intra- and interspecific level which in turn supported the opportunity to use dosymbionts gene as a marker to identify host evolutionary history (Liu et al. 2013).

Out of 11 species, 7 were identified in *Cyclocheilichthys apogon*, 8 in *Labiobarbus festivus*, and 7 in *Pristipomoides filamentosus* (Table 5). Bacterial abundance was significantly varies among fish species. Overall mean bacterial abundance (CFU/g) were higher in *C. apogon* (6.68×10^3) compare to *L. festivus* (5.12×10^3) and *P. filamentosus* (5.20×10^3) (Fig. 3). Overall, highest bacterial abundances were observed in fish gut (6.62×10^3), followed by fish gill (5.78×10^3) and fish skin (4.60×10^3) (Fig. 4). However, these results are different depending on fish species (Fig. 5). However, this trend was only observed in *C. apogon*. In the case of *L. festivus* and *P. filamentosus*, abundances of bacteria were lower in skin compared to gill and gut. Abundances of bacteria in gill and gut were statistically similar in both *L. festivus* and *P. filamentosus.* There is no previous study in the Kuantan River comparing the bacterial abundance in various parts of the fish body. However, Alikunhi et al. (2017) observed similar trend (gut > gill > skin) of bacterial abundance in 13 species of commonly consumed fish collected from three places in the Jeddah region Saudi Arabia, although, their observed bacterial abundance in various places of fish body were higher compare to the present observation. Many factors affect the abundance of bacteria in various organs of fish body. Fish normally contaminate with bacteria from water, sediment, and food; hence, their abundance in fish is largely influenced by the condition of their habitat. The observed bacterial abundances in various organs of various fishes were under the safe limit of aerobic bacteria count ($5 \times 10^5 - 5 \times 10^7$ cfu/g for fresh fish) (ICMSF 1986). However, regular monitoring of bacterial load in the Kuantan River fish is suggested to keep track of any potential health risks to fish consumers.

Fig. 1 (continued) together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Kimura 2 parameter method (Kimura 1980) and are in the units of the number of base substitutions per site. The analysis involved 62 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 881 positions in the final dataset

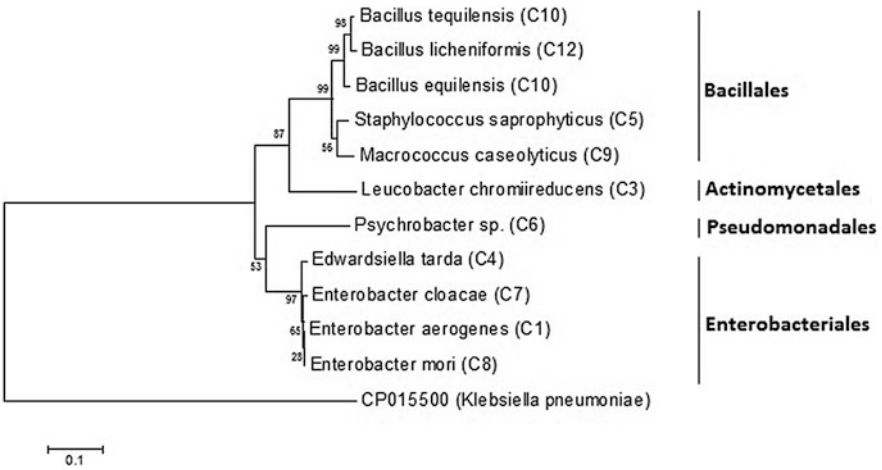


Fig. 2 Neighbor-joining dendrogram depicting the estimated phylogenetic relationships among unknown (indicating C1, C2, etc. in parentheses) species based on 16S rRNA GenBank. The optimal tree with the sum of branch length = 1.75303059 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Kimura 2 parameter method (Kimura 1980) and are in the units of the number of base substitutions per site. The analysis involved 12 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 893 positions in the final dataset

Table 3 Pair-wise genetic distances (Kimura 2 parameter) of 11 Kuantan River fish bacteria based on 16S rRNA sequences

	C1	C8	C7	C4	C6	C10	C12	C13	C5	C9	C3
C1											
C8	0.001										
C7	0.006	0.007									
C4	0.014	0.015	0.015								
C6	0.225	0.227	0.230	0.228							
C10	0.280	0.282	0.287	0.281	0.345						
C12	0.286	0.287	0.292	0.286	0.34	0.016					
C13	0.274	0.276	0.282	0.276	0.325	0.026	0.031				
C5	0.275	0.277	0.282	0.282	0.32	0.082	0.780	0.074			
C9	0.277	0.279	0.284	0.282	0.325	0.070	0.065	0.065	0.054		
C3	0.282	0.283	0.280	0.292	0.332	0.244	0.242	0.232	0.236	0.239	

C1, *Enterobacter aerogenes*; C3, *Leucobacter chromiireducens*; C4, *Edwardsiella tarda*; C5, *Staphylococcus saprophyticus*; C6, *Psychrobacter* sp.; C7, *Enterobacter cloacae*; C8, *Enterobacter mori*; C9, *Macrococcus caseolyticus*; C10, *Bacillus tequilensis*; C12, *Bacillus licheniformis*; C13, *Bacillus pumilus*

Table 4 Pair-wise genetic distances (Kimura 2 parameter) among four groups (Fig. 2) of Kuantan River fish bacteria based on 16S rRNA sequences

Group	Enterobacteriales	Pseudomonadales	Bacillales	Actinomycetales
Enterobacteriales				
Pseudomonadales	0.229			
Bacillales	0.282	0.330		
Actinomycetales	0.285	0.331	0.235	

Table 5 List of bacteria identified from skin, gill, and gut of kerisi, chemparus, and kawan collected from the Kuantan River

	Cyclocheilichthys apogon	*Labiobarbus festivus*	*Pristipomoides filamentosus*
Skin			
Enterobacter aerogenes	×	×	×
Edwardsiella tarda			×
Leucobacter chromiireducens	×	×	
Micrococcus caseolyticus			×
Staphylococcus saprophyticus	×	×	×
Gill			
Enterobacter aerogenes	×		×
Edwardsiella tarda	×		×
Leucobacter chromiireducens		×	
Micrococcus caseolyticus	×		×
Psychrobacter sp.	×		×
Staphylococcus saprophyticus		×	×
Gut			
Edwardsiella tarda	×		
Bacillus pumilus	×	×	
Bacillus licheniformis		×	
Bacillus tequilensis		×	
Enterobacter mori			
Enterobacter cloacae			×
Micrococcus caseolyticus	×	×	
Psychrobacter sp.	×	×	×
Staphylococcus saprophyticus			×

"×" indicates presence

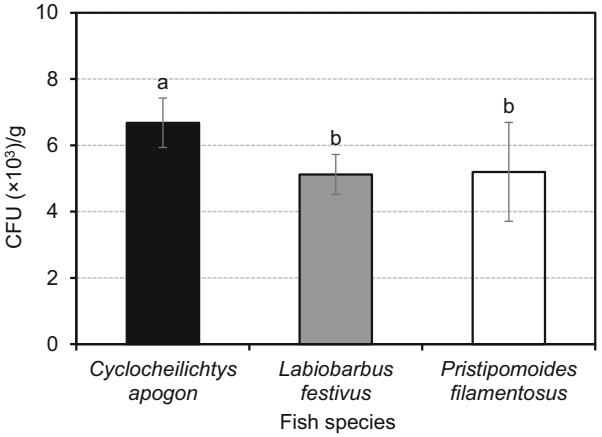


Fig. 3 Mean (bar: ±SD) number of cultivable bacteria (CFU/g) in chemparus, kawan, and kerisi. Mean with no letters in common is significantly different ($P < 0.05$)

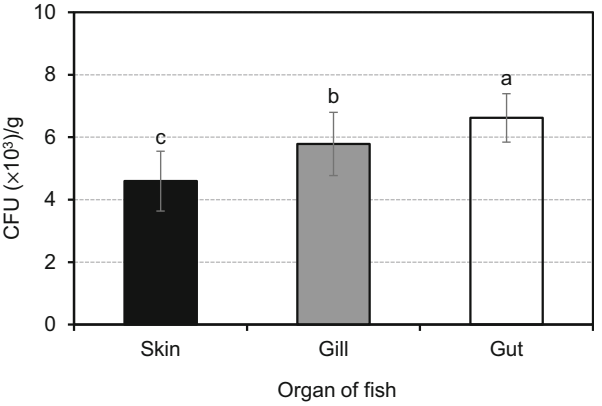


Fig. 4 Mean (bar: ±SD) number of cultivable bacteria (CFU/g) in skin, gill, and gut of fishes. Mean with no letters in common is significantly different ($P < 0.05$)

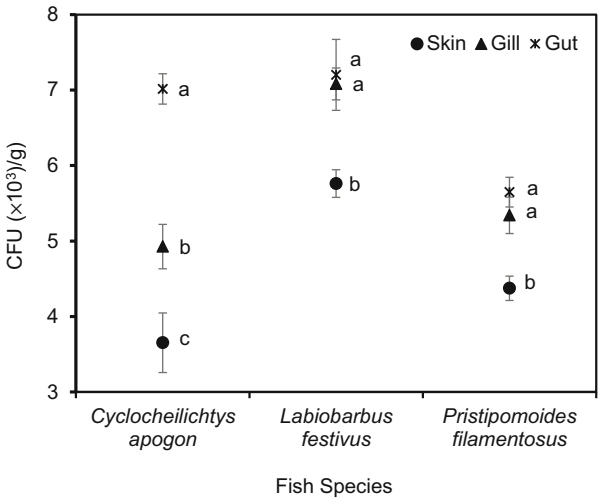


Fig. 5 Interaction between fish species and fish organ on number (per g) of cultivable bacteria

References

Alikunhi NM, Batang ZB, AlJahdali HA, Aziz MAM, Al-Suwailem AM (2017) Culture dependent bacteria in commercial fishes: qualitative assessment and molecular identification using 16S rRNA gene sequencing. Saudi J Biol Sci. https://doi.org/10.1016/j.sjbs.2016.05.017

Alperi A, Figueras MJ, Inza I, Martínez-Murcia AJ (2008) Analysis of 16S rRNA gene mutations in a subset of *Aeromonas* strains and their impact in species delineation. Int Microbiol 11:185–194

Alzahrani OM (2015) Molecular identification of Aeromonas spp. collected from Taif Province of Saudi Arabia. Int J Sci Res 4:2359–2362

Austin B, Stuckey LF, Robertson PAW, Effendi I, Griffith DRW (1995) A probiotic strain of *Vibrio alginolyticus* effective in reducing diseases caused by *Aeromonas salmonicida*, *Vibrio anguillarum* and *Vibrio ordalii*. J Fish Dis 18:93–96

Begerow D, Nilsson H, Unterseher M, Maier W (2010) Current state and perspectives of fungal DNA barcoding and rapid identification procedures. Appl Microbiol Biotechnol 87:99–108

Bhattacharya M, Sharma AR, Patra BC, Sharma G, Seo EM, Nam JS, Chakraborty C, Lee SS (2016) DNA barcoding to fishes: current status and future directions. Mitochondrial DNA A DNA Mapp Seq Anal 27:2744–2752

Bienfang PK, DeFelice SV, Laws EA, Brand LE, Bidigare RR, Christensen S, Trapido-Rosenthal H, Hemscheidt TK, DJ MG Jr, Anderson DM, Solo-Gabriele HM, Boehm AB, Backer LC (2011) Prominent human health impacts from several marine microbes: history, ecology, and public health implications. Int J Microbiol 2011:1–15

Brady C, Cleenwerck I, Venter S, Vancanneyt M, Swings J, Coutinho T (2008) Phylogeny and identification of *Pantoea* species associated with plants, humans and the natural environment based on multilocus sequence analysis (MLSA). Syst Appl Microbiol 31:447–460

Cottrell M, Waidner LA, Yu L, Kirchman DL (2005) Bacterial diversity of metagenomic and PCR libraries from the Delaware river. Environ Microbiol 7:1883–1895

de Sousa JS, Silva-Souza AT (2001) Bacterial community associated with fish and water from Congonhas River, Sertaneja, Parana, Brazil. Braz Arch Biol Technol 44:373–381

Deleetoile A, Decre D, Courant S, Passet V, Audo J, Grimont P, Arlet G, Brisse S (2009) Phylogeny and identification of Pantoea species and typing of Pantoea agglomerans strains by multilocus gene sequencing. J Clin Microbiol 47:300–331

Drancourt M, Bollet C, Carta A, Rousselier P (2001) Phylogenetic analyses of *Klebsiella* species delineate *Klebsiella* and *Raoultella* gen. nov., with description of *Raoultella ornithinolytica* comb. nov., *Raoultella terrigena* comb. nov. and *Raoultella planticola* comb. nov. Int J Syst Evol Microbiol 51:925–932

Dunbar J, Barns SM, Ticknor LO, Kuske CR (2002) Empirical and theoretical bacterial diversity in four Arizona soils. Appl Environ Microbiol 68:3035–3045

El-Shafai SA, Gijzen HJ, Nasr FA, El-Gohary FA (2004) Microbial quality of tilapia reared in fecal contaminated ponds. Environ Res 95:231–238

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Figueras MJ, Alperi A, Beaz-Hidalgo R, Stackebrandt E, Brambilla E, Monera A, Martínez-Murcia AJ (2011) *Aeromonas rivuli* sp. nov., isolated from the upstream region of a karst water rivulet in Germany. Int J Syst Evol Microbiol 61:242–248

Garcıa-Lopez I, Otero A, Garcıa-Lopez M-L, Santos JA (2004) Molecular and phenotypic characterization of nonmotile Gram-negative bacteria associated with spoilage of freshwater fish. J Appl Microbiol 96:878–886

Han HJ, Kim DY, Kim WS, Kim CS, Jung SJ, Oh MJ, Kim DH (2011) Atypical *Aeromonas salmonicida* infection in the black rockfish, *Sebastes schlegeli* Hilgendorf, in Korea. J Fish Dis 34:47–55

Hasan MR, Hecht T, De Silva SS, Tacon AGJ (2007) Study and analysis of feeds and fertilizers for sustainable aquaculture development. FAO Fisheries Technical Paper. No. 497. Rome, 510 p

Hauben L, Moore ERB, Vauterinl L, Steenackers M, Mergaert J, Verdonck L, Swings J (1998) Phylogenetic position of phytopathogens within the Enterobacteriaceae. Syst Appl Microbiol 21:384–397

Hebert PDN, Gregory TR (2005) The promise of DNA barcoding for taxonomy. Syst Biol 54:852–859

Hickerson MJ, Meyer CP, Moritz C (2006) DNA barcoding will often fail to discover new animal species over broad parameter space. Syst Biol 55:729–739

Hidalgo RB, Figueras MJ (2012) Molecular detection and characterization of Furunculosis and other *Aeromonas* fish infections. In: Carvalho ED, David GS, Silva RJ (eds) Health and environment in aquaculture. InTech, Rijeka, Croatia, pp 97–132

Hugenholtz P (2002) Exploring prokaryotic diversity in the genomic era. Genome Biol 3: Reviews 0003

International Commission on Microbiological Specifications for Foods (ICMSF) (1986) Microorganisms in foods 2: sampling for microbiological analysis: principles and specific applications, 2nd edn. Blackwell Scientific, Oxford, p 190

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kwon H-S, Yang E-H, Yeon S-W, Kang B-H, Kim T-Y (2004) Rapid identification of probiotic Lactobacillus species by multiplex PCR using species-specific primers based on the region extending from 16S rRNA through 23S rRNA. FEMS Microbiol Lett 239:267–275

Links MG, Dumonceaux TJ, Hemmingsen SM, Hill JE (2012) The Chaperonin-60 universal target is a barcode for bacteria that enables de novo assembly of metagenomic sequence data. PLoS ONE 7(11):e49755

Liu L, Huang X, Zhang R, Jiang L, Qiao G (2013) Phylogenetic congruence between Mollitrichosiphum (Aphididae: Greenideinae) and Buchnera indicates insect-bacteria parallel evolution. Syst Entomol 38:81–92

Ma B, Hibbing ME, Kim H-S, Reedy RM, Yedidia I, Breuer J, Breuer J, Glasner JD, Perna NT, Kelman A, Charkowski AO (2007) Host range and molecular phylogenies of the soft rot enterobacterial genera *Pectobacterium* and *Dickeya*. Bacteriology 97:1150–1163

Maruyama A, Honda D, Yamamoto H, Kitamura K, Higashihara T (2000) Phylogenetic analysis of psychrophilic bacteria isolated from the Japan Trench, including a description of the deep-sea species Psychrobacter pacificensis sp nov. Int J Syst Evol Microbiol 50:835–846

Novotny L, Dvorska L, Lorencova A, Beran B, Pavlik I (2004) Fish: a potential source of bacterial pathogens for human beings. Vet Med Czech 49:343–358

Olafsen JA (2001) Interactions between fish larvae and bacteria in marine aquaculture. Aquaculture 200:223–247

Purty RS, Chatterjee S (2016) DNA barcoding: an effective technique in molecular taxonomy. Austin J Biotechnol Bioeng 3:1–10

Rahman MM (2015a) Role of common carp (*Cyprinus carpio*) in aquaculture production systems. Front Life Sci 8:399–410

Rahman MM (2015b) Effects of co-cultured common carp on nutrients and food web dynamics in rohu aquaculture ponds. Aquac Environ Interact 6:223–232

Rahman MM, Meyer CG (2009) Effects of food type on diel behaviours of common carp *Cyprinus carpio* L. in simulated aquaculture pond conditions. J Fish Biol 74:2269–2278

Rahman MM, Verdegem MCJ (2007) Multi-species fishpond and nutrients balance. In: ven der Zijpp AJ, AJA V, Tri LQ, ven Mensvoort MEF, Bosma RH, Beveridge MCM (eds) Fishponds in farming systems. Wageningen Academic Publishers, Wageningen

Rahman MM, Verdegem MCJ, Wahab MA (2008a) Effects of tilapia (*Oreochromis nilotica* L.) addition and artificial feeding on water quality, and fish growth and production in rohu-common carp bi-culture ponds. Aquac Res 39:1579−1587

Rahman MM, Verdegem M, Nagelkerke L, Wahab MA, Milstein A, Verreth J (2008b) Effects of common carp *Cyprinus carpio* (L.) and feed addition in rohu *Labeo rohita* (Hamilton) ponds on nutrient partitioning among fish, plankton and benthos. Aquac Res 39:85−95

Rahman MM, Hossain MY, Jo Q, Kim SK, Ohtomi J, Meyer CG (2009) Ontogenetic shift in dietary preference and low dietary overlap in rohu (*Labeo rohita* Hamilton) and common carp (*Cyprinus carpio* L.) in semi-intensive polyculture ponds. Ichthyol Res 56:28–36

Rastogi G, Sani RK (2011) Molecular techniques to assess microbial community structure, function, and dynamics in the environment. In: Ahmad I, Ahmad F, Pichtel J (eds) Microbes and microbial technology: agricultural and environmental applications. Springer, New York, pp 29–57

Robertson PAW, O-Dowd C, Burrells C, Williams P, Austin B (2000) Use of *Carnobacterium* sp. as a probiotic for Atlantic salmon (*Salmo salar* L.) and rainbow trout (*Oncorhynchus mykiss*, Walbaum). Aquaculture 185:235–243

Schmid M, Iversen C, Gontia I, Stephan R, Hofmann A, Hartmann A, Jha B, Eberl L, Riedel K, Lehner A (2009) Evidence for a plant-associated natural habitat for *Cronobacter* spp. Res Microbiol 160:608–614

Sebastiao FA, Furlan LR, Hashimoto DT, Pilarski F (2015) Identification of bacterial fish pathogens in Brazil by direct colony PCR and 16S rRNA gene sequencing. Adv Microbiol 5:409–424

Smith EM, Prairie YT (2004) Bacterial metabolism and growth efficiency in lakes: the importance of phosphorus availability. Limnol Oceanogr 49:137–147

Sprber C, Mendrock U, Swiderski J, Lang E, Stackebrandt E (1998) The phylogenetic position of *Serratia, Buttiauxella* and some other genera of the family *Enterobacteriaceae*. Int J Syst Bacteriol 49:1431–1438

Stahlhut JK, Gibbs J, Sheffield CS, Smith MA, Packer L (2012) *Wolbachia* (Rickettsiales) infections and bee (Apoidea) barcoding: a response to Gerth et al. Syst Biodivers 10:395–401

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

Van Horn DJ, Sinsabaugh RL, Takacs-Vesbach CD, Mitchell KR, Dahm CN (2016) Response of heterotrophic stream biofilm communities to a gradient of resources. Aquat Microb Ecol 64:149–161

Wang Y, Jiang X, Xu Z, Ying C, Yu W, Xiao Y (2016) Identification of *Raoultella terrigena* as a rare causative agent of subungual abscess based on 16S rRNA and housekeeping gene sequencing. Can J Infect Dis Med Microbiol (in press). https://doi.org/10.1155/2016/3879635

Zhang Y, Qiu S (2015) Examining phylogenetic relationships of *Erwinia* and *Pantoea* species using whole genome sequence data. Antonie Van Leeuwenhoek 108:1037–1046

DNA Barcoding of Ichthyoplankton and Juvenile Fishes of a Tropical River in Malaysia

B. Akbar John, Hassan I. Sheikh, K. C. A. Jalal, B. Y. Kamaruzzaman, H. Sanower, M. Nur Hanisah, M. H. Rahman, and M. Rozihan

Abstract Taxonomic identification of early larval stages of fishes using conventional morphological keys is extremely laborious due to the overlapping characters shared between genetically closer species. Especially, species-level differentiation during their ontological development is challenging due to paucity of information on their diagnostic features. In the present study, we aimed to use universal DNA barcoding technology to identify ichthyoplankton and juvenile fishes of a tropical river (Kuantan River in Pahang) in Malaysia. This sampling station was chosen in order to check the distribution of juvenile and fish ichthyoplankton samples after the recent massive flood encountered in East Peninsular Malaysia during 2014. We adopted mitochondrial cytochrome oxidase C subunit 1 gene sequencing to identify fish samples. A total of 28 species from 15 families and 5 orders were identified successfully to the species level from the total of 58 DNA barcodes. Unlike the previous report, the most dominant fishes in this study belong to the Cyprinidae family followed by Toxotidae, Ambassidae, and Eleotridae. We admit that the modified bubu light trap method adopted for larval collection in this study has its limitation to attract larvae which had negative phototactic behavior (i.e., Ariidae fishes). Phylogenetic and BLAST analysis showed accuracy of species identification with high bootstrap and percentage similarity value, respectively. Results in this study confirmed the efficiency of universal DNA barcode technology in species-level delimitation of morphologically cryptic species identification. The data

B. Akbar John (✉) · K. C. A. Jalal
INOCEM Research Station (IRS), Kulliyyah of Science, International Islamic University Malaysia (IIUM), Kuantan, Pahang, Malaysia

H. I. Sheikh · H. Sanower · M. Nur Hanisah
Department of Biotechnology, Kulliyyah of Science, International Islamic University Malaysia (IIUM), Kuantan, Pahang, Malaysia

B. Y. Kamaruzzaman · M. H. Rahman
Department of Marine Science, Kulliyyah of Science, International Islamic University Malaysia (IIUM), Kuantan, Pahang, Malaysia

M. Rozihan
Department of Aquaculture, University Putra Malaysia (UPM), Selangor, Malaysia

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_23

presented in this study is valuable for analyzing post-flood effect on fish distribution in tropical river and implementing plans for future fishery resource management in Kuantan River, Pahang, Malaysia.

Keywords DNA barcoding · COX1 gene · Fish larvae · Ichthyoplankton · Kuantan River

1 Introduction

Ichthyoplankton and early juvenile stages of fishes are the most sensitive forms, and their distribution is altered by the ambient environmental condition. Their assemblage is directly proportional to spawning strategies of their brooders (especially in the estuary) and driven by water currents and mass. In tropics, rivers, in general, have direct connection with the sea, and their physical and topological parameters are severely altered during massive flood. For instance, recent flood incident in East Peninsular Malaysia from December 2014 to mid-January 2015 caused massive damage to the livelihood and altered sensitive species distribution in almost all major rivers. From a regional perspective, this flood had significant impact on fish distribution and species richness that has influenced the livelihood of fishery folk in Kuantan River skirts besides planning various management measures for future sustainable fishery. On the other hand, in taxonomic point of view, accurate identification of ichthyoplankton and early juvenile stages of fishes is challenging even to an expert taxonomist. It is due to the lack of information on their early ontological development, overlapping and continuously changing in morphological characters, and, thus, empirical study of larval ecology is currently limited (Choat et al. 1993; Nakatani et al. 1998). Most species identification is based on the data availability on their adult subjects (Ko et al. 2013; Pegg et al. 2006) and many morpho-diagnostic characters or poorly developed in their larval forms especially in their egg stage (Baumgartner et al. 2004; Nakatani 2001).

The utility of standardized molecular marker in accurately identifying morphologically cryptic species using universal barcode gene (Cytochrome oxidase C subunit 1) has proven to be a promising method in species discrimination (Kochzius et al. 2010). Though the reliability of single gene sequencing to identify metazoan species is argued by many researchers (Cowan et al. 2006; Decru et al. 2016; DeSalle et al. 2005; John et al. 2016; Meier et al. 2006; Vences et al. 2005), the method is still promising in identifying early life stages, and adults of many taxon (Bucklin et al. 2011; Hajibabaei et al. 2006; Hausmann et al. 2016; John et al. 2016; Spasojevic et al. 2016; Ward et al. 2005). Ko et al. (2013) have suggested that accuracy of fish larval identification using morphological, body pigmentation and meristic count is possible with only 3–30% of samples compared to the 100% accuracy of DNA barcoding technique (Ratnasingham and Hebert 2007). The method is rapid and accurate and helps in differentiating morphologically cryptic species and avoids

inconsistency in species-level identification of sensitive forms (Frantine-Silva et al. 2015; Hubert et al. 2015; Ko et al. 2013; Overdyk et al. 2016).

DNA barcoding technique has been successfully used in accurate identification of early larval forms (pre- and post-flexion stages), eggs, and pre-juvenile stages and adopted in various ecological studies in recent years (Frantine-Silva et al. 2015; Gleason and Burton 2012; Pegg et al. 2006; Valdez-Moreno et al. 2010). Many studies have even argued that molecular taxonomy is the only promising tool for identifying eggs to their corresponding species (Lewis et al. 2016). Nevertheless, the paper addresses the utility of DNA barcoding in ichthyoplankton and juvenile fish identification and their post-flood distribution in Kuantan River, Pahang, Malaysia. The data presented in this study is crucial for implementing various fishery management measures for future sustainable fishery in tropical riverine system exposed to severe flood.

2 Materials and Methods

2.1 Sample Collection

Ichthyoplankton and juvenile fish samples were collected from Kuantan River and its immediate tributaries during ebb tidal cycle (Fig. 1). Due to the bottom topology and physical condition of Kuantan riverbed, such as high turbidity and uneven water depth in Kuantan River (varied from 4 m to 13 m) besides inefficiency of plankton sampling net (mesh size 500 μ), a modified bubu light trap (Fig. 2) was developed in this study and used during sampling from April to December 2015. Modified bubu light trap used in this study is 5 × 4 × 3 ft (L × W × H) size having a conical-shaped opening in one side toward the interior section of the trap. The skeleton is made up of bamboo or cylindrical wood which in turn covered completely by 2 mm thickness stainless steel wire meshes with the mesh diameter of 2.5 cm^2. An underwater light was placed in transparent plastic container, and the lid was sealed with commercially available silicone gel and parafilm to ensure maximum light emission. An anchor was tied at the bottom of the cage to make sure the cage does not wash away during water current. The complete set up (modified bubu light trap) was cover by plastic window mesh sheet (mesh size $<$ 1 mm) (Fig. 2). The net was deployed underwater at the depth of 4 m in sampling stations for 16 h overnight. All samples were stored in 70% ethanol for laboratory identification.

2.2 Sample Preparation and Identification

Ichthyoplankton and juvenile fish samples were measured under DinoCapture 2.0 portable microscope. Samples were identified morphologically to the lowest possible taxon using standard references and taxonomically classified based on Marques et al. (2006). Size classes of larvae were tabulated and represented in Mean ± SD.

Fig. 1 Location of the sampling sites at Kuantan River

2.3 DNA Barcoding

Total genomic DNA was extracted using Geneaid DNA Tissue Isolation Kit^{TM} following manufacturer instruction. NanoDrop 2000 UV-Vis spectrophotometer was used to quantify the total genomic DNA. Approximately 650 bp of partial cytochrome oxidase C subunit 1 gene was amplified using standard thermal cycler in polymerase chain reaction (PCR) using the primer set (Fish F1, 5-′-GGTCAACAAATCATAAAGATATTGG-3′, and Fish R1, 5-′-TAAACTTCAGGGTGACCAAAAAATCA-3′). The amplification reactions were performed in a final volume including 6.3 µl of molecular grade water, 1.0 µl of 10× PCR buffer, 0.5 µl of $MgCl_2$ (50 mM), 0.3 µl of each primer (10 mM), 0.5 µl of dNTPs (10 mM), 0.1 µl of Bioline Taq polymerase, and 1 µl of template DNA. The PCR condition includes hot start with 94 °C for 1 min, 5 cycles of 94 °C for 30 s, annealing at 45 °C for 40 s, and extension at 72 °C for 1 min and 35 cycles of 94 °C for 30 s, 51 °C for 40 s, and final extension at 72 °C for 10 min. Final amplicon were purified by adding 0.5 µl Illustra Exo-Star 1-Step PCR Clean Up Kit (Thermo Fisher Scientific, Waltham, MA) and gel eluded in 2% China Agarose gel and the photographed using Gel imager Under UV light for future reference. DNA sequencing was carried at First Base Sdn.Bhd using ABI sequencer.

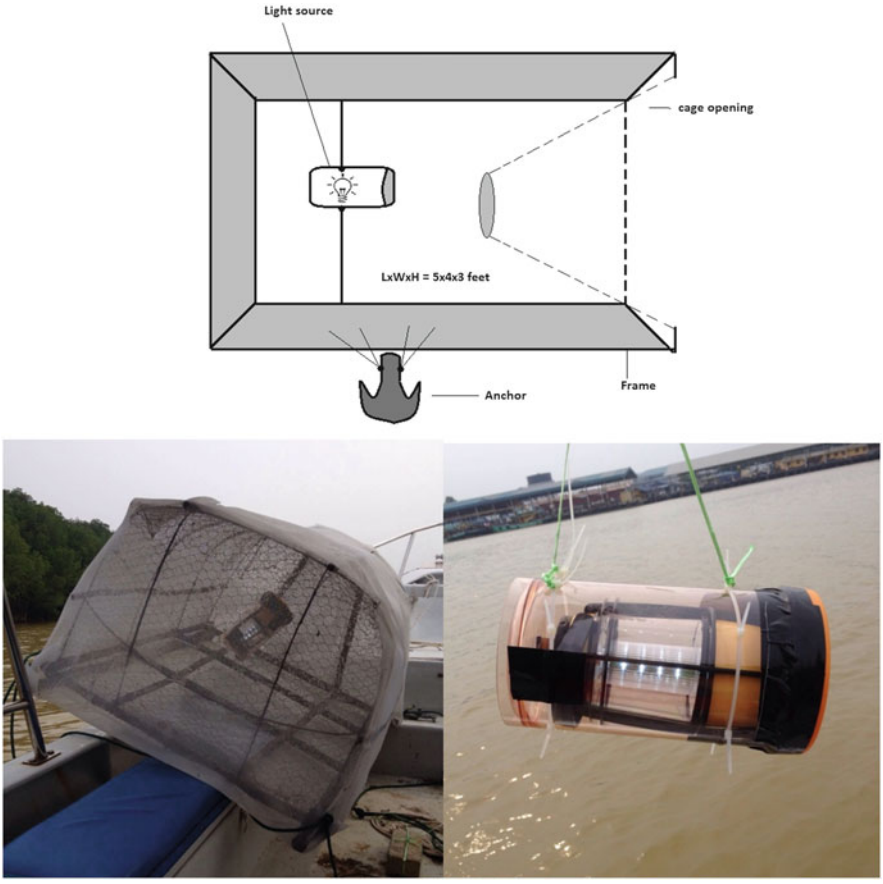


Fig. 2 Modified bubu light trap and the prepared internal light source used to attract the ichthyoplankton during sampling

2.4 Data Analysis

DNA sequences were trimmed using sequence scanner software v2.0. The trimmed sequences with no insertion, deletion, and stop codon were subjected to Barcode of Life Database BOLD (http://www.boldsystems.org/index.php/IDS_OpenIdEngine) and NCBI BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch) analysis for species identification and cross-examined with morphological identification. Multiple sequence alignment (MSA) was performed in Clustal X 2.0.6v using default settings (Larkin et al. 2007). Nucleotide composition was computed using Bioedit v7.2.5 (Hall 1999). Species identification reliability was validated in evolutionary tree constructed using neighbor-joining (NJ) tree method using Kimura 2-parameter (K2P) model using default condition with 1000 bootstrap replicates. Genetic distance were calculated using K2P distance matrix using MEGA 6.0V (Tamura et al. 2013).

3 Results and Discussion

During the sampling period, 372 larval/juvenile fish samples were captured in modified bubu light trap for further molecular identification. Samples with full appendages were selected for DNA isolation and for further downstream studied. Overall, 28 species of ichthyoplankton/juvenile fishes from 15 families and 5 orders were identified using DNA barcoding COI gene. Unlike previous report on adult fish distribution in Kuantan River by Jalal et al. (2012) who observed dominant fishes belong to the families such as Ariidae, Lactaridae, and Lutjanidae, the dominant families in this study belong to Cyprinidae, Toxotidae, Ambassidae, and Eleotridae with the percentage abundance of 35%, 24%, 18%, and 11%, respectively (Table 1). We have adopted modified light trap method to collect the ichthyoplankton and a juvenile fish due to many technical constrains to use standard bongo net for sampling. It includes mainly the bottom topology of riverbed and constant changes in river depth (3–10 m) among the immediate tributaries. Other constrains include high turbidity and organic debris on the water column which hindered the collection of live and full form of larvae (without any appendage lost). Though, this light trap method is effective to attract many species of larvae, the fishes exhibiting negative phototactic behavior such as Ariidae larvae, mouth brooders were not trapped in the net (Mukai et al. 2008; Satoh et al. 2017). The advantages of using light trap includes (1) the collection of undamaged larval form for downstream application without other species DNA contamination, (2) efficiency of light trap over plankton net with 0.5 mm mesh size, and (3) samplings in areas where turbidity and water depth are manor limiting factors.

3.1 Database Analysis

DNA barcode generated from 58 samples (>650 bp) were subjected to BOLD and NCBI BLAST analysis independently. A match of 95–100% was found in 72% of test sequences, and 7% samples showed >90% similarity, while 21% samples were identified with more than 85% similarity value. Almost 47 out of 58 sequences exhibited higher similarity percentage with best match to nearest neighbor indicating ambiguity of ~19% of barcode assignment to their accurate species. Overall, 81% of all barcodes showed <2% divergence from their possible nearest neighbors. From the barcode generated, dominant Cyprinidae family represented with eight species (*Poropuntius smedleyi, Barbonymus schwanenfeldii*, *B. gonionotus*, *Neolissochilus stracheyi*, *Barbodes binotatus*, *Rasbora sumatrana*, *R. dusonensis*, *and R. trilineata*) followed by Toxotidae (*Toxotes chatareus*, *T. jaculatrix*), Ambassidae (*Ambassis marianus*, *A. commersoni*, *and Parambassis siamensis*), *and* Eleotridae (*Butis gymnopomus*, *Prionobutis dasyrhynchus*).

Table 1 Ichthyoplankton and juvenile fishes identified in this study using universal DNA barcode (Cytochrome oxidase C subunit 1 gene)

Family	Species identified	IUC N red list status	BOLD/ BLAST similarity percentage	Kuantan River (N 03° 47′ 04.2″ E 103° 19′ 05.1″)	Belat River (N 03° 46′ 15.7″ E 103° 17′ 21.7″)	Panching River (N 03° 47′ 4.65″ E 103° 08′ 58.5″)	Riau River (N 03° 51′ 51.6″ E 103° 13′ 32.4″)	Previous study (Jalal et al. 2012)
Ambassidae	*Ambassis marianus*	Least concern	92	*				NF
Ambassidae	*Ambassis commersoni*	Least concern	85		*			NF
Ambassidae	*Parambassis siamensis*	Least concern	96			*		NF
Leiognathidae	*Secutor ruconius*	Not evaluated	100	*	*			NF
Leiognathidae	*Photopectoralis bindus*	Not evaluated	99		*			NF
Leiognathidae	*Leiognathus equulus*	Least concern	99	*				NF
Lutjanidae	*Lutjanus johnii*	Not evaluated	84	*				*
Lutjanidae	*Lutjanus russellii*	Not evaluated	100	*				*
Eleotridae	*Butis gymnopomus*	Not evaluated	100	*				NF
Eleotridae	*Prionobutis dasyrhynchus*	Not evaluated	99	*			*	NF
Toxotidae	*Toxotes chatareus*	Not evaluated	100		*			*
Toxotidae	*Toxotes jaculatrix*	Least concern	99		*			*

(continued)

Table 1 (continued)

Family	Species identified	IUC N red list status	BOLD/ BLAST similarity percentage	Kuantan River (N 03° 47′ 04.2″ E 103° 19′ 05.1″)	Belat River (N 03° 46′ 15.7″ E 103° 17′ 21.7″)	Panching River (N 03° 47′ 4.65″ E 103° 08′ 58.5″)	Riau River (N 03° 51′ 51.6″ E 103° 13′ 32.4″)	Previous study (Jalal et al. 2012)
Cyprinidae	*Poropuntius smedleyi*	Not evaluated	99			*		NF
Cyprinidae	*Barbonymus schwanenfeldii*	Least concern	100				*	NF
Cyprinidae	*Barbonymus gonionotus*	Least concern	99				*	NF
Cyprinidae	*Neolissochilus stracheyi*	Least concern	98			*		NF
Cyprinidae	*Barbodes binotatus*	Least concern	99		*			NF
Cyprinidae	*Rasbora sumatrana*	Not evaluated	99			*		NF
Cyprinidae	*Rasbora dusonensis*	Not evaluated	99				*	NF
Cyprinidae	*Rasbora trilineata*	Least concern	97				*	NF
Megalopidae	*Megalops cyprinoides*	Data deficient	100	*				*
Gobiidae	*Bathygobius laddi*	Not evaluated	86	*				NF
Cichlidae	*Oreochromis niloticus*	Not evaluated	84		*			*
Percichthyidae	*Gadopsis marmoratus*	Not evaluated	85		*			NF

Monodactylidae	*Monodactylus argenteus*	Not evaluated	90		*			NF
Scatophagidae	*Scatophagus argus*	Least concern	99		*			NF
Zenarchopteridae	*Dermogenys pusilla*	Not evaluated	99		*			NF
Hemiramphidae	*Hemirhamphodon pogonognathus*	Least concern	100			*		NF
Balitoridae	*Balitora kwangsiensis*	Least concern	83				*	NF

Note: * represents the inhabiting species in each sampling station. Species identified in this study was compared with previous study where adult fishes were sampled using gill net (Jalal et al. 2012). NF represent corresponding species in the row "Not Found" during previous study

3.2 Phylogenetic Analysis

Phylogenetic tree constructed using NJ method and K2P distance model showed clear segregation of identified taxa to their corresponding species. Though there are some species mismatches observed in the tree [in the case of Lutjanidae (*Lutjanus russellii*) and Monodactylidae (*Monodactylus argenteus*)], high bootstrap value in internal nods >98–100% signified reliability of constructed phylogram with low distance between samples within the clusters formed by the species (Fig. 3). The observed nucleotide composition for all species is as follows: A (25.8%), G (18%), T (28.8), and C (27.4). Overall pairwise distance was observed to be 0.337 with an average P distance of 0.243. Genetic distance was hierarchically increased from lowest possible taxon to the higher taxonomic level. Intraspecific genetic distance was observed between 0 and 2.56% compared to interspecies genetic distance (6.08–14.18%) when the proposed genetic distance threshold value of 2% for intraspecific and 6.8% for congeneric species is used (Frantine-Silva et al. 2015; Pereira et al. 2013). Nucleotide substitution pattern observed among the test organisms (including all larvae/juveniles barcoded) clearly showed that transitional substitutions are very common in the gene sequence than transversional substitutions with transition/transversion bias value of 1.54 (Table 2).

Many attempts were made to analyze ichthyoplankton assemblage changes and its composition in marine (Ko et al. 2013; Pegg et al. 2006; Valdez-Moreno et al. 2010) and inland waters of Brazil (Frantine-Silva et al. 2015) and Australia (Loh et al. 2014). However, studies on molecular identification of ichthyoplankton and early life stages from tropical river is still limited. In this study, we present first report on the efficiency of molecular identification technique to accurately discriminate species sampled from one of the major river in East Peninsular Malaysia. Malaysia has faced severe northeast monsoonal hit from mid-December 2014 to January 2015 with at least >60% above the normal precipitation rate. This in turn increased uncontrolled discharge of organic and inorganic substances in to the river stream which are believed to be the limiting factor of sensitive organisms such as juvenile and ichthyoplankton fishes. In order to accurately identify immediate successive (i.e., post-flood) juvenile and ichthyoplankton samples, we have adopted COI gene sequence similarity in comparison with public databanks as proposed by Ratnasingham and Hebert (2007) and Ward et al. (2009). Our analysis clearly showed efficiency of universal barcode gene (COI) in species identification, although percentage similarity and genetic distance approaches introduce some bias (Taylor and Harris 2012). Unlike the previous study by Ko et al. (2013) and Frantine-Silva et al. (2015) who observed 85% and 99% of detection efficiency of COI gene sequencing to the species level, our study depicted slightly lower ~81% and 100% species and genus level identification, respectively. It might be probably due to the availability of less reference sequences submitted especially on juvenile fishes sampled from tropical rivers. In fact, almost all samples identified in this study generally inhabit in tropical freshwater ecosystem. Although, DNA level identification is accurate and rapid approach, there is an urgent need of conventional taxonomic

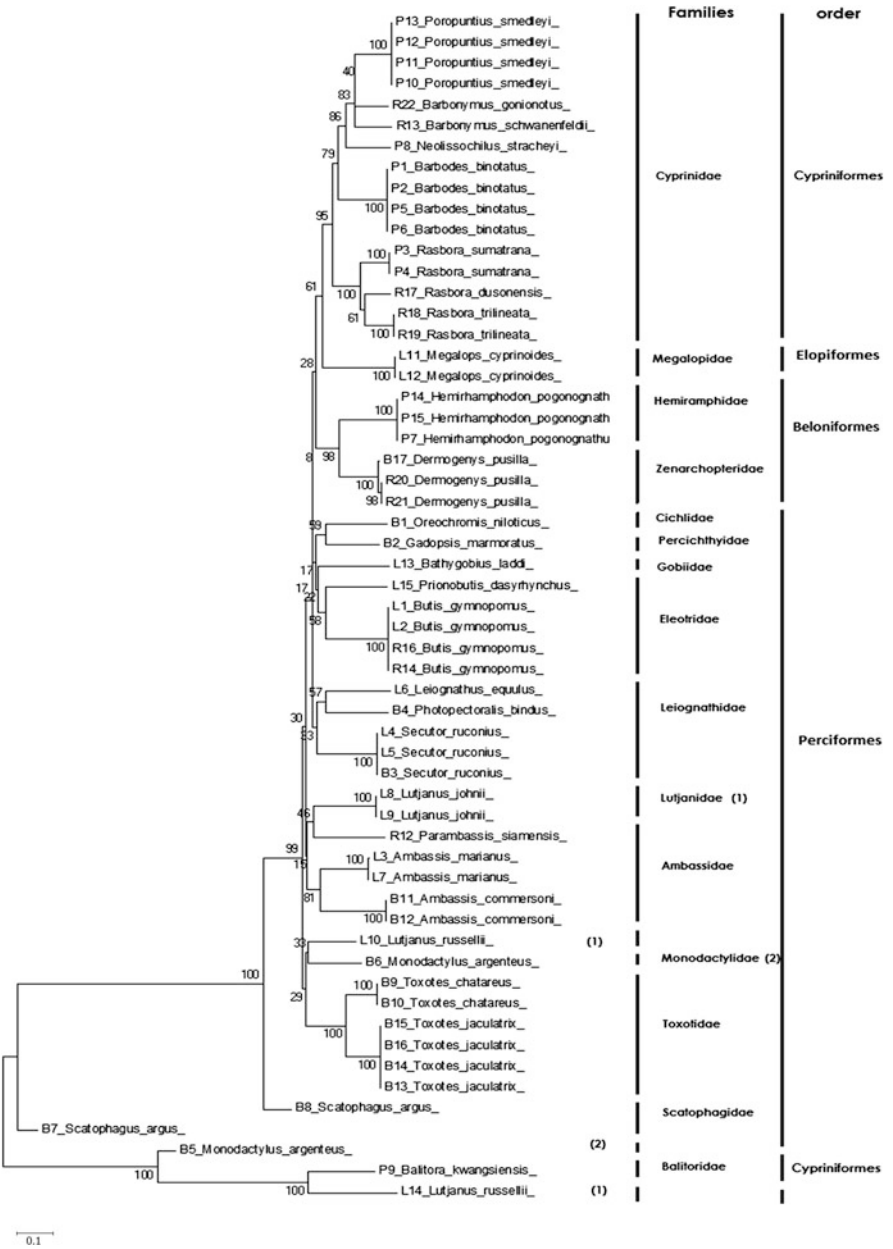


Fig. 3 Evolutionary relationships of taxa collected in this study. The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei 1987). The optimal tree with the sum of branch length = 3.94044074 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were

Table 2 Maximum likelihood estimate of substitution matrix and pattern of nucleotide substitution observed among the test organisms

	A	T/U	C	G
A	–	*4.91*	*4.91*	**15.17**
T/U	*4.91*	–	**15.17**	*4.91*
C	*4.91*	**15.17**	–	*4.91*
G	**15.17**	*4.91*	*4.91*	–

Note: Each entry is the probability of substitution (r) from one base (row) to another base (column). Substitution pattern and rates were estimated under the Kimura (1980) 2-parameter model. Rates of different transitional substitutions are shown in **bold** and those of transversional substitutions are shown in *italics*. Relative values of instantaneous r should be considered when evaluating them. For simplicity, sum of r values is made equal to 100. The nucleotide frequencies are A = 25.00%, T/U = 25.00%, C = 25.00%, and G = 25.00%. For estimating ML values, a tree topology was automatically computed. The maximum log-likelihood for this computation was −10,101.892. The analysis involved 57 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 553 positions in the final dataset. Evolutionary analyses were conducted in MEGA6

identification keys to be developed for ichthyoplankton samples for cross-examining results from DNA study. Several models have been proposed to match unknown sequences to the known species. For example, in coalescent models, mutational models, distance-based models, and phylogenetic and mixed phylogenetic-coalescent models (Brown et al. 2012; David et al. 2012; Munch et al. 2008a, b; Pons et al. 2006; Puillandre et al. 2012) that in turn require phylogenetic diversity and well-sampled intraspecific parameters (Ratnasingham and Hebert 2013), similar data is still unavailable for tropical fishes, though some attempts were shown promising results (Cooper et al. 2009; Hubert et al. 2011; Westneat and Alfaro 2005). Hence, these approaches cannot be used unless the comprehensive DNA barcode library is established. However, the present study using percentage similarity approach to identify generated barcodes showed initial promising results, though the limited availability of morphological characters.

Fig. 3 (continued) computed using the Kimura 2-parameter method (Kimura 1980) and are in the units of the number of base substitutions per site. The analysis involved 57 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 553 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013)

4 Conclusion

In conclusion, DNA barcoding of ichthyoplankton and juvenile fishes using COI gene sequencing showed promising, rapid, and accurate identification of samples to their corresponding species level. The post-flood effect on ichthyoplankton species distribution in Kuantan River showed apparent shift in species composition which might have positive impact on fishery in near future. Though, an unavailability of corresponding species DNA barcode data in public databanks is one of the limiting factor for species delimitation, almost 80% of the samples were identified with higher percentage similarity (>99%). It is worth mentioning that the identification of 28 species of juvenile forms of fishes from Kuantan River compared to the previous report on conventional taxonomy (19 species) would increase our understanding on greater diversity of species at least in Malaysian in riverine system. Constant monitoring in fish composition and distribution would pave a way in various fishery management enforcement practices toward sustainable fishery in East Peninsular Malaysia particularly in Kuantan River.

Acknowledgment The research work was sponsored under special flood Fundamental Research Grant Scheme (FRGS), Ministry of Science and Technology, Malaysia.

References

Baumgartner G, Nakatani K, Gomes LC, Bialetzki A, Sanches PV (2004) Identification of spawning sites and natural nurseries of fishes in the upper Paraná River, Brazil. Environ Biol Fish 71(2):115–125

Brown SD, Collins RA, Boyer S, Lefort MC, Malumbres-Olarte J, Vink CJ, Cruickshank RH (2012) Spider: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. Mol Ecol Resour 12(3):562–565

Bucklin A, Steinke D, Blanco-Bercial L (2011) DNA barcoding of marine metazoa. Annu Rev Mar Sci 3:471–508

Choat HJ, Doherty PP, Kerrigan BB, Leis JJ (1993) A comparison of towed nets, purse seine and light-aggregation devices for sampling of larvae and pelagic stages of coral reef fishes. Fish Bull 91:195–209

Cooper WJ, Smith LL, Westneat MW (2009) Exploring the radiation of a diverse reef fish family: phylogenetics of the damselfishes (Pomacentridae), with new classifications based on molecular analyses of all genera. Mol Phylogenet Evol 52(1):1–16

Cowan RS, Chase MW, Kress WJ, Savolainen V (2006) 300,000 species to identify: problems, progress, and prospects in DNA barcoding of land plants. Taxon 55(3):611–616

David O, Larédo C, Leblois R, Schaeffer B, Vergne N (2012) Coalescent-based DNA barcoding: multilocus analysis and robustness. J Comput Biol 19(3):271–278

Decru E, Moelants T, De Gelas K, Vreven E, Verheyen E, Snoeks J (2016) Taxonomic challenges in freshwater fishes: a mismatch between morphology and DNA barcoding in fish of the north-eastern part of the Congo basin. Mol Ecol Resour 16(1):342–352

DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. Philos Trans R Soc B Biol Sci 360(1462):1905–1916

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Frantine-Silva W, Sofia S, Orsi M, Almeida F (2015) DNA barcoding of freshwater ichthyoplankton in the Neotropics as a tool for ecological monitoring. Mol Ecol Resour 15(5): 1226–1237

Gleason LU, Burton RS (2012) High-throughput molecular identification of fish eggs using multiplex suspension bead arrays. Mol Ecol Resour 12(1):57–66

Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PD (2006) DNA barcodes distinguish species of tropical Lepidoptera. Proc Natl Acad Sci U S A 103(4):968–971

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Paper presented at the Nucleic acids symposium series

Hausmann A, Miller SE, Holloway JD, deWaard JR, Pollock D, Prosser SW, Hebert PD (2016) Calibrating the taxonomy of a megadiverse insect family: 3000 DNA barcodes from geometrid type specimens (Lepidoptera, Geometridae) 1. Genome 59(9):671–684

Hubert N, Paradis E, Bruggemann H, Planes S (2011) Community assembly and diversification in Indo-Pacific coral reef fishes. Ecol Evol 1(3):229–277

Hubert N, Espiau B, Meyer C, Planes S (2015) Identifying the ichthyoplankton of a coral reef using DNA barcodes. Mol Ecol Resour 15(1):57–67

Jalal K, Azfar MA, John BA, Kamaruzzaman Y, Shahbudin S (2012) Diversity and community composition of fishes in tropical estuary Pahang Malaysia. Pak J Zool 44(1):181–187

John BA, Sheikh HI, Jalal K, Zaleha K, Kamaruzzaman B (2016) Revised phylogeny of extant Xiphosurans (Horseshoe Crabs) DNA barcoding in marine perspectives. Springer, Switzerland, pp 113–130

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16(2):111–120

Ko H-L, Wang Y-T, Chiu T-S, Lee M-A, Leu M-Y, Chang K-Z, Shao K-T (2013) Evaluating the accuracy of morphological identification of larval fishes by applying DNA barcoding. PLoS One 8(1):e53451

Kochzius M, Seidel C, Antoniou A, Botla SK, Campo D, Cariani A, Hjörleifsdottir S (2010) Identifying fishes through DNA barcodes and microarrays. PLoS One 5(9):e12620

Larkin MA, Blackshields G, Brown N, Chenna R, McGettigan PA, McWilliam H, Lopez R (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23(21):2947–2948

Lewis LA, Richardson DE, Zakharov EV, Hanner R (2016) Integrating DNA barcoding of fish eggs into ichthyoplankton monitoring programs. Fish Bull 114(2):153–166

Loh W, Bond P, Ashton K, Roberts D, Tibbetts I (2014) DNA barcoding of freshwater fishes and the development of a quantitative qPCR assay for the species-specific detection and quantification of fish larvae from plankton samples. J Fish Biol 85(2):307–328

Marques SC, Azeiteiro UM, Marques JC, Neto JM, Pardal MÂ (2006) Zooplankton and ichthyoplankton communities in a temperate estuary: spatial and temporal patterns. J Plankton Res 28(3):297–312

Meier R, Shiyang K, Vaidya G, Ng PK (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Syst Biol 55(5):715–728

Mukai Y, Tuzan AD, Lim LS, Wahid N, Raehanah S, Senoo S (2008) Development of sensory organs in larvae of African catfish Clarias gariepinus. J Fish Biol 73(7):1648–1661

Munch K, Boomsma W, Huelsenbeck JP, Willerslev E, Nielsen R (2008a) Statistical assignment of DNA sequences using Bayesian phylogenetics. Syst Biol 57(5):750–757

Munch K, Boomsma W, Willerslev E, Nielsen R (2008b) Fast phylogenetic DNA barcoding. Philos Trans R Soc Lond B Biol Sci 363(1512):3997–4002

Nakatani K (2001) Ovos e larvas de peixes de água doce: desenvolvimento e manual de identificação: Eletrobrás. Uem 2001:il

Nakatani K, Baumgartner G, Latini JD (1998) Morphological description of larvae of the mapara Hypophthalmus edentatus (Spix)(Osteichthyes, Hypophthalmidae) in the Itaipu reservoir (Parana River, Brazil). Revista Brasileira de Zoologia 15(3):687–696

Overdyk LM, Holm E, Crawford SS, Hanner RH (2016) Increased taxonomic resolution of Laurentian Great Lakes ichthyoplankton through DNA barcoding: a case study comparison against visual identification of larval fishes from Stokes Bay, Lake Huron. J Great Lakes Res 42(4):812–818

Pegg GG, Sinclair B, Briskey L, Aspden WJ (2006) MtDNA barcode identification of fish larvae in the southern Great Barrier Reef–Australia. Sci Mar 70(S2):7–12

Pereira LH, Hanner R, Foresti F, Oliveira C (2013) Can DNA barcoding accurately discriminate megadiverse Neotropical freshwater fish fauna? BMC Genet 14(1):20

Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Vogler AP (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Syst Biol 55(4):595–609

Puillandre N, Lambert A, Brouillet S, Achaz G (2012) ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Mol Ecol 21(8):1864–1877

Ratnasingham S, Hebert PD (2007) BOLD: the barcode of life data system. Mol Ecol Notes 7(3): 355–364. http://www.barcodinglife.org

Ratnasingham S, Hebert PD (2013) A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. PLoS One 8(7):e66213

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4):406–425

Satoh S, Tanoue H, Ruitton S, Mohri M, Komatsu T (2017) Morphological and behavioral ontogeny in larval and early juvenile discus fish Symphysodon aequifasciatus. Ichthyol Res 64(1): 37–44

Spasojevic T, Kropf C, Nentwig W, Lasut L (2016) Combining morphology, DNA sequences, and morphometrics: revising closely related species in the orb-weaving spider genus Araniella (Araneae, Araneidae). Zootaxa 4111(4):448–470

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30(12):2725–2729

Taylor H, Harris W (2012) An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. Mol Ecol Resour 12(3):377–388

Valdez-Moreno M, Vásquez-Yeomans L, Elías-Gutiérrez M, Ivanova NV, Hebert PD (2010) Using DNA barcodes to connect adults and early life stages of marine fishes from the Yucatan Peninsula, Mexico: potential in fisheries management. Mar Freshw Res 61(6):655–671

Vences M, Thomas M, Bonett RM, Vieites DR (2005) Deciphering amphibian diversity through DNA barcoding: chances and challenges. Philos Trans R Soc Lond B Biol Sci 360(1462): 1859–1868

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PD (2005) DNA barcoding Australia's fish species. Philos Trans R Soc Lond B Biol Sci 360(1462):1847–1857

Ward RD, Hanner R, Hebert PD (2009) The campaign to DNA barcode all fishes, FISH-BOL. J Fish Biol 74(2):329–356

Westneat MW, Alfaro ME (2005) Phylogenetic relationships and evolutionary history of the reef fish family Labridae. Mol Phylogenet Evol 36(2):370–390

Genetic Variation of Wild and Hatchery Populations of Climbing Perch, *Anabas testudineus* (Bloch, 1792), in Peninsular Malaysia

Ahmad Azfar Mohamed, Siti Waznah Abdurahman, and Akbar John

Abstract In order to increase the production of *Anabas testudineus*, selective breeding can be implemented to produce better performing stock. To initiate a successful selective breeding program, genetic variation is very important. This study was aimed to aid selective breeding of *A. testudineus* in Malaysia. Cytochrome b gene was selected as a molecular marker to investigate the current status of genetic variation in hatchery and wild populations of *A. testudineus*. Based on cytochrome *b* analysis, all existing hatchery populations in Malaysia belonged to a single haplotype. In contrast, five distinct haplotypes were discovered based on genotypic data of wild *A. testudineus* populations in Malaysia. Overall, this study recommended to introduce new genetic material from wild population to the synthetic base population to increase the genetic variation of the populations of *A. testudineus*. All of the information gathered is crucial to develop a sustainable selective breeding for *A. testudineus* and to improve quality and production of this species.

Keywords *Anabas testudineus* · Genetic · Hatchery · Cytochrome *b* · Climbing perch

A. A. Mohamed (✉)
INOCEM Research Station, Kulliyyah of Science, International Islamic University Malaysia, Kuantan, Pahang, Malaysia

Department of Marine Science, Kulliyyah of Science, International Islamic University Malaysia, Kuantan, Pahang, Malaysia
e-mail: azfar@iium.edu.my

S. W. Abdurahman
Department of Marine Science, Kulliyyah of Science, International Islamic University Malaysia, Kuantan, Pahang, Malaysia

A. John
INOCEM Research Station, Kulliyyah of Science, International Islamic University Malaysia, Kuantan, Pahang, Malaysia

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_24

1 Introduction

Anabas testudineus (Bloch, 1972) belongs to suborder Anabantoidei and is commonly known as climbing perch. Locally *A. testudineus* is known as ikan puyu or ikan betok. According to Ambak et al. (2010), it is indigenous to Malaysia, and only one species is recorded. This species can be found in sluggish, standing, or even stagnant water with dense vegetation, and it has a wide range of geographical distribution. It can be found mostly in rivers, canals, lakes, ponds, swamps, and paddy fields (Bungas et al. 2013). *Anabas testudineus* are geographically distributed in waters of India, Pakistan, Bangladesh, Nepal, China, Myanmar, Thailand, Cambodia, Malaysia, the Philippines, Indonesia, Singapore, and Sri Lanka (Zalina et al. 2012).

Anabas testudineus is highly valued and economically important food fish in Asia, because of its good taste, flesh quality and value, and prolonged freshness out of water (Rahman et al. 2013). Production of this species was increasing rapidly over the years, and the number of farms producing this fish is growing, due to the high demand (Hidayat and Senanan 2010).

Genetic variation is very important for a population; by maintaining high genetic variation and diversity in a population, it may increase long-term viability of a population to adapt to environmental changes and increase fitness and selection (Filgueira et al. 2010). However, in hatchery population the loss of genetic variation is a common phenomenon caused by inbreeding and genetic drift. For this species, there is still no official report or clear indication of genetic abnormalities caused by loss of genetic variation. To avoid this phenomenon occurring to this species, it is very important to evaluate the genetic variations in hatchery and wild populations of this species. Therefore, this study was carried out in order to evaluate the genetic variations of wild and hatchery populations of *A. testudineus* in Peninsular Malaysia using universal mitochondrial DNA marker of cytochrome *b*.

2 Methodology

2.1 Sample Collection

The samples were collected at different sites along Peninsular Malaysia from May 2011 to October 2013, and samples from Vietnam and Thailand were collected from December 2012 to June 2013, respectively. The wild samples were bought from fishermen and markets, but most of the samples were caught using gill net and angling techniques. For the hatchery samples, samples were bought from hatcheries listed by DOF (Department of Fisheries) directory as main *A. testudineus* captive producer. All the collected samples were brought to the laboratory for further morphological and molecular analyses.

2.2 Sample Analyses

Total DNA of the samples collected was extracted using a commercial DNA extraction kit (BioTeke Corporation, Beijing, China). Approximately 30–40 mg tissue of the samples was extracted by following the kit manufacturer protocol. Purified DNA concentration was estimated using spectrophotometer. For immediate use, purified DNA was maintained at 4 °C and kept at −20 °C for storage.

For the isolation of partial cytochrome b mitochondrial DNA, approximately 50–100 ng of the DNA templates were amplified using universal cyt *b* gene primer, namely, *cyt B1*, for forward reaction (5′-CCA TCC AAC ATC TCA GCA TGA TGA AA-3′), and *cyt B2*, for reverse reaction (5′-GCC CCT CAG AAT GAT ATT TGT CCT CA-3′) following Carr and Marshall (1991). The PCR amplification was performed in 25.0 μL PCR mixture consisted of 16.6 μL of deionized water, 2.5 μL of PCR buffer (1×), 1.0 μL of $MgCl_2$ (2.0 mM), 0.5 μL of each dNTP (0.2 mM), 0.6 μL of each forward and reverse primers (25 pmol), 0.2 μL of Taq polymerase (1U), and 3.0 μL of DNA template. Amplification condition was consisted of a pre-denaturation at 94 °C for 3 min, 30 cycles of denaturation at 95 °C for 1 min, annealing step for 1 min at 50 °C and primer extension step at 72 °C for 2 min, and final extension step at 72 °C for 10 min.

DNA amplification fragments were separated by electrophoresis at 80 V on 1.5% agarose gel (added GelRed as staining agent) with Tris-borate-EDTA buffer for 90 min (Sambrook et al. 1989). Agarose gel was viewed using UV illumination (GelDoc), and digital photograph was used to estimate the fragment length by comparing it with 1 kb ladder (λ DNA/Hind III, Vivantis, Malaysia).

Nucleic acid sequences were analyzed using BLASTN (Altschul et al. 1990) searches at the National Center for Biotechnology Information (NCBI), and they were aligned using ClustalW software (Thompson 1991). The Chromas (McCarthy 1997), BioEdit (Hall 1999), and MEGA 4.0 (Tamura et al. 2007) software were used in nucleotide alignment and editing sequences. The phylogenetic relationships among samples were calculated based on maximum parsimony (MP) and neighbor joining (NJ) in MEGA 4.0. The MP calculates relationship based on characteristic, while NJ used distance to predict the relationship. The DnaSP (Rozas et al. 2003) and Arlequin 3.1 (Excoffier et al. 2005) software were used to analyze genetic diversity from aligned data set.

3 Results

A total of 224 samples from 18 population consisted of wild and hatchery samples were successfully amplified for cyt *b* gene. After alignment and editing, a total length of 307 bp from cyt *b* gene for all the samples was obtained. From these 18 populations, 12 of them were from wild and the other 6 were from hatchery populations including hatcheries from Thailand and Vietnam.

From the obtained 307 bp cyt *b* gene, a total of 38 variable/polymorphic sites were observed, which covering 12.38% of total nucleotide discovered, while another 87.62% of total nucleotide was monomorphic/conserved nucleotide sites (Table 1). From 38 polymorphic sites, all 38 sites were found parsimony-informative site, which means that the variable sites occur with a minimum frequency of 2. In another word, there is no singleton variable site found within 38 variable/polymorphic sites. In further analysis of the variable/polymorphic sites, one parsimony-informative site with three variants was identified at nucleotide position 49.

Overall, 38 substitutions were found among six haplotypes where transition/transversion rate ratio for purines (K_1) was 17.246 and for pyrimidines (K_2) was 8.608 and the overall transition/transversion bias (R) calculated was 6.044. From the analyzed samples, the mean nucleotide frequency was 22.86% A, 29.44% T, 30.31% C, and 17.38% G. From mean nucleotide frequency, there was an anti-G bias within data set where there was less C and G (47.7%), compared to A and T (52.3%); this characteristic is indicating mitochondrial gene (Tingga and Abdullah 2011). From six haplotypes discovered, the first polymorphic site occurs at position 1 with substitution of C to T in Hap03, and the last polymorphic site found in the analysis of the total nucleotide was at nucleotide position 292 with substitution of A to G in Hap03. Overall, the nucleotide substitution within single site was only composing of two nucleotides, and the highest number of substitution observed was in haplotypes three.

From overall samples, the highest number of samples was contributed by Hap01 with total percentage of 60.35%, Hap02 was contributing 18.06% of total number of sample making it the second highest, and the lowest number of sample was recorded by Hap03 with total percentage of 1.32%. The rest of the samples were belonging to Hap05, Hap06, and Hap04 with total percentage of 10.13%, 5.73%, and 4.41%, respectively. Most of the haplotype found was shared by different populations except Hap03, Hap04, and Hap06 that are unique to its own population. The haplotype distribution explained is clearly illustrated in Table 2.

The Hap01 was appeared to be the most frequent haplotype to be observed among all populations. Among wild populations, Hap01 was observed within all population except in Johor, and the highest number of sample was observed in Kedah population. Within Malaysian hatchery populations, Hap01 was present in all population, and the highest number of sample was observed in Kelantan and Selangor populations. The Hap01 was observed as dominant in hatchery populations as it is only absent in Vietnam hatchery population.

Among all haplotypes, three haplotypes were found as population-specific haplotype as it is only available in a specific population. The Hap03 was only observed in Kedah wild population, Hap04 was only found in Selangor wild population, and Hap06 was exclusive to Thailand hatchery population. Between these population-specific haplotypes, the highest frequency observed was Hap06. Further observation on this unique haplotype revealed that Hap03 and Hap04 belong to wild population and only Hap06 belong to hatchery population.

From all populations, Selangor wild population contain the highest number of haplotypes. Within Selangor population, three different haplotypes were observed,

Table 1 Haplotype variations and nucleotide polymorphic sites encoded by six haplotype within hatchery and wild *A. testudineus* populations generated by cytochrome *b* mtDNA gene

No.	Nucleotide variable sites	Haplotype					
		Hap01	Hap02	Hap03	Hap04	Hap05	Hap06
1	1	C	–	T	–	–	–
2	13	C	–	T	–	–	–
3	16	A	–	T	–	–	–
4	19	G	C	G	C	C	C
5	30	G	–	A	–	–	–
6	37	G	–	C	–	–	–
7	41	T	–	A	–	–	C
8	50	T	–	C	–	–	–
9	61	A	–	C	–	–	–
10	70	T	C	C	–	–	–
11	71	G	A	A	A	G	A
12	73	G	–	T	–	–	–
13	76	A	–	G	–	–	–
14	109	T	C	T	C	C	C
15	115	T	–	C	–	–	–
16	160	C	–	–	–	–	T
17	172	T	–	C	–	–	–
18	178	C	–	T	–	–	–
19	196	C	–	T	–	–	–
20	205	A	G	G	G	G	G
21	214	T	–	C	–	–	–
22	217	C	–	T	–	–	–
23	223	T	–	C	–	–	–
24	226	C	–	T	–	–	–
25	232	C	–	T	–	–	–
26	244	G	A	A	A	A	A
27	253	G	–	A	–	–	–
28	256	T	C	–	–	–	–
29	263	T	–	C	–	–	–
30	265	A	–	G	–	–	–
31	269	C	–	T	–	–	–
32	271	A	–	G	–	–	–
33	274	T	–	C	–	–	C
34	277	A	–	G	–	–	–
35	280	G	–	A	–	–	–
36	289	C	–	T	–	–	–
37	292	A	–	G	–	–	–
Haplotype frequency		0.621	0.130	0.034	0.130	0.011	0.073

Table 2 Haplotype distributions across all hatchery and wild *A. testudineus* populations generated by cytochrome *b* mtDNA gene

		Hap01	Hap02	Hap03	Hap04	Hap05	Hap06
Wild	Kelantan	0.035	0.070				
	Kedah	0.079		0.013			
	Selangor	0.044			0.044	0.013	
	Johor		0.110				
Hatchery	Kelantan	1.101					
	Kedah	0.097					
	Selangor	0.101					
	Johor	0.093					
	Vietnam					0.088	
	Thailand	0.053					0.057
	%	60.35	18.06	1.32	4.41	10.13	5.73

Table 3 Pairwise genetic distance among haplotypes of *A. testudineus* inferred from cytochrome *b* mtDNA gene

No.	Haplotype	1	2	3	4	5	6
1	Hap01	–					
2	Hap02	0.023	–				
3	Hap03	$0.119^{\uparrow}$	0.115	–			
4	Hap04	0.017	0.007	0.115	–		
5	Hap05	0.013	0.010	0.119	$0.003^{\downarrow}$	–	
6	Hap06	0.030	0.020	0.119	0.013	0.017	–

namely, Hap01, Hap03, and Hap04. On the other hand, majority of the population was only observed with single haplotype. The Hap01 was observed as sole haplotype in all hatchery populations except Thailand. Johor wild population and Vietnam hatchery population also possessed single haplotype which is Hap02 and Hap05, respectively.

Wild populations of *A. testudineus* were observed with multiple haplotypes. Among isolated six haplotypes, distinguished five haplotypes were found within the wild populations. Each of the wild population contains multiple haplotypes, whereas only Johor wild population dominated by single haplotype. Contrarily, most of the hatchery populations were observed with single haplotype except Thailand population. Even though, among all hatchery populations, only three different haplotypes were observed.

The pairwise genetic distance calculated the number of nucleotide differences; Tamura-Nei model was used to calculate pairwise genetic distance among haplotypes. In general, this analysis shows that the genetic divergent among samples haplotypes. A clear genetic differentiation was observed among all haplotypes as in Table 3.

The analysis of pairwise genetic distance among haplotypes showed that the highest value of pairwise genetic distance was 11.9%, while the lowest was 0.3%.

The highest divergent among haplotypes was observed between Hap03 and Hap01, Hap05 and Hap03, and also Hap06 and Hap03. The lowest divergent among haplotypes was observed between Hap05 and Hap04. Overall, from pairwise genetic distance among haplotypes, the most divergent haplotype observed was Hap03. The pairwise genetic distance value was highest between Hap03 and all other haplotypes. However, according to Hebert et al. (2003), there is divergent distance matrix value between species that are greater than 3%. But he is suggesting to employing threshold value for species diagnosis as 3%. Therefore, Hap03 was not considered as haplotype. Thus, it was not included in later analysis.

Pairwise genetic distance among populations was calculated using Tamura-Nei model as in Table 4. Among all populations, the wild populations demonstrate genetic differentiation. Among hatchery populations, the genetic differentiation was not observed except for Thailand and Vietnam populations. The pairwise genetic distance recorded was ranging from 0% to 3.0% of genetic differentiation between populations. Overall, the value of pairwise genetic distance was higher among wild populations compared to hatchery populations. Among wild populations, the genetic differentiation was ranging from 1.0% to 3.0%. The highest genetic distance was observed between Johor and Selangor (3.0%) populations. The lowest genetic distance observed was 1.0% between Kedah and Kelantan populations.

Between wild and hatchery populations, genetic differentiation observed between populations with pairwise genetic distance value was from 0% to 2.6%. The highest genetic distance value observed was between Vietnam hatchery population and Selangor wild population. There is no genetic distance observed between Kedah wild population and all hatchery populations except Thailand and Vietnam hatchery populations.

Neighbor-joining and maximum parsimony trees were constructed using K2+I (Kimura 2 parameter) to analyze phylogenetic relationships of cyt *b* mtDNA genes in the data set. In this analysis, *Helostoma temminckii* (kissing gourami) and *Osphronemus goramy* (giant gourami) were used as out-group. Neighbor-joining and maximum parsimony analyses showed different tree topologies for phylogenetic relationships. The phylogenetic trees were constructed using 15 sequences from each population representing each haplotype to show relationships between haplotypes and populations.

The neighbor-joining and maximum parsimony trees are as in Figs. 1 and 2, respectively. Both figures showed that the highest variation within haplotype was contributed by Kelantan and Johor wild populations. It is also suggested that the variation between haplotypes was highly contributed by wild populations compared to hatchery populations. From the dendrogram, hatchery populations have been clustered into a single clade. However, Thailand and Vietnam hatchery populations were closely clustered together with low bootstrap confidence together with Selangor wild population. From the dendrogram, wild populations were scattered among different clusters. A distinct cluster was observed, containing wild populations with low bootstrap confidence.

Table 5 shows the genetic diversity within populations. The table showed that only wild populations showed genetic diversity within populations. There is no

Table 4 Pairwise genetic distance among wild and hatchery populations of *A. testudineus* inferred from cytochrome *b* mtDNA gene

	No.	Population	1	2	3	4	5	6	7	8	9
Wild	1	Kelantan									
	2	Kedah	0.010								
	3	Selangor	0.022	0.023							
	4	Johor	0.022	0.023	0.030[↑]						
Hatchery	5	Kelantan	0.010	0.000	0.023	0.023					
	6	Kedah	0.010	0.000	0.023	0.023	0.000				
	7	Selangor	0.010	0.000	0.023	0.023	0.000	0.000			
	8	Johor	0.010	0.000	0.023	0.023	0.000	0.000	0.000		
	9	Vietnam	0.004	0.017	0.026	0.007	0.017	0.017	0.017	0.017	
	10	Thailand	0.006	0.008	0.021	0.014	0.008	0.008	0.008	0.008	0.007

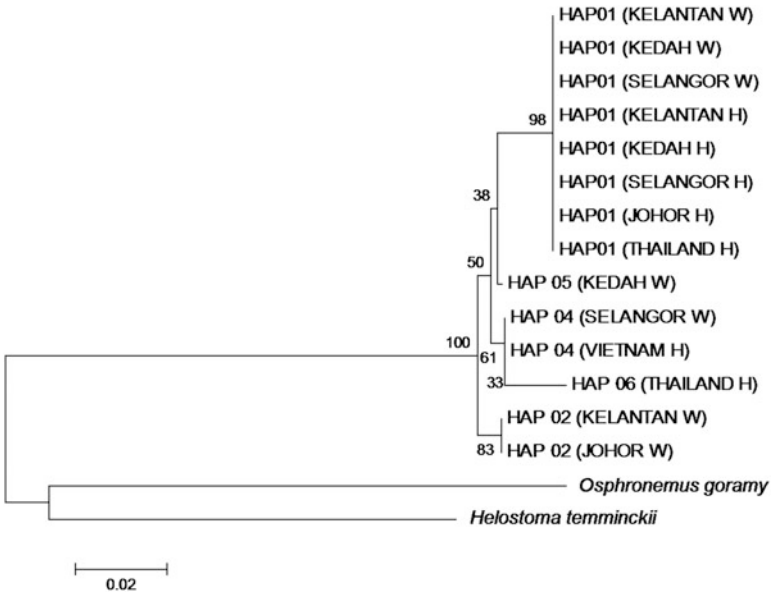


Fig. 1 Phylogenetic relationship among different populations of *A. testudineus* based on cytochrome *b* mtDNA gene. The phylogenetic tree was inferred using the neighbor-joining method. The percentage of replicate trees in which the associated haplotype clustered together in the bootstrap test (1000 replicates) are shown next to the branches (only the bootstrap values >50% are shown)

genetic variation observed within hatchery populations. Among wild populations, Kelantan and Selangor populations showed higher genetic diversity within each population. The result of genetic diversity within populations demonstrates different situations among hatchery populations. There was almost no genetic diversity observed within hatchery populations. However, Thailand hatchery population does show genetic differentiation within its population.

The value of nucleotide diversity within populations varied differently, with the Pi value ranging from 0 to 0.05678. The levels of nucleotide diversity within populations demonstrate that wild populations have higher value of nucleotide diversity compared to hatchery populations. Among wild populations, Selangor wild population shows highest nucleotide diversity within population with Pi value of 0.05678 and Kelantan wild population with the Pi value 0.01057. Among hatchery populations, only Thailand hatchery population shows nucleotide diversity within population. However, there was monomorphic event observed among all populations. All of the hatchery populations were monomorphic except Thailand population; among wild populations only Johor and Kedah wild populations were monomorphic.

Within each population, the number of haplotype was between one and three haplotypes within populations. There were five haplotypes observed within wild populations, and only three haplotypes were observed within hatchery populations. Haplotype diversity calculated within wild and hatchery populations revealed that wild

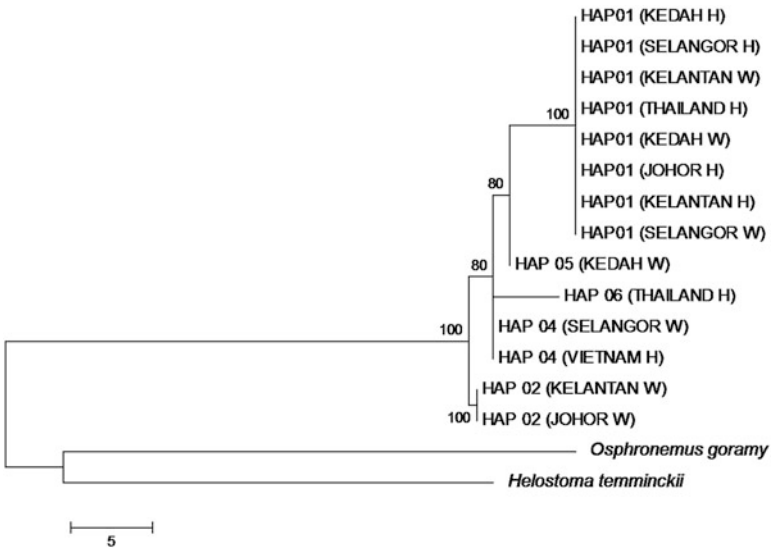


Fig. 2 Phylogenetic relationship among different populations of *A. testudineus* based on cytochrome *b* mtDNA gene. The phylogenetic tree was inferred using the maximum parsimony method. The percentage of replicate trees in which the associated haplotype clustered together in the bootstrap test (1000 replicates) are shown next to the branches (only the bootstrap values >50% are shown)

Table 5 Population genetic diversity of *A. testudineus* inferred from cytochrome *b* gene

Population		V	H	H_d	K	Pi
Wild	Kelantan	7	2	0.46377	3.24638	0.01057
	Kedah	0	1	0.0000	0.0000	0.0000
	Selangor	35	3	0.63241	17.43083	0.05678
	Johor	0	1	0.0000	0.0000	0.0000
Hatchery	Kelantan	0	1	0.0000	0.0000	0.0000
	Kedah	0	1	0.0000	0.0000	0.0000
	Selangor	0	1	0.0000	0.0000	0.0000
	Johor	0	1	0.0000	0.0000	0.0000
	Vietnam	0	1	0.0000	0.0000	0.0000
	Thailand	9	2	0.5200	4.6800	0.01524
Wild		36	5	0.64914	9.08789	0.02960
Hatchery		9	3	0.40321	2.57625	0.00839

Note: Number of variable site (V), number of haplotype (H), haplotype diversity (H_d), average number of differences (K), and nucleotide diversity (Pi)

populations have higher haplotype diversity compared to hatchery populations with the value of 0.64914 and 0.40321, respectively. Among wild populations, haplotype diversity was ranging from 0 to 0.63241. The highest diversity was observed within Selangor wild population, and the haplotype diversity with zero value was recorded by

monomorphic populations (Johor and Kedah wild populations). However, only Thailand hatchery population shows haplotype diversity (0.5200), and the rest of hatchery populations were monomorphic.

4 Discussion

The low genetic diversity and differentiation among hatchery populations were observed throughout the study. From phylogenetic analyses, the trees constructed have grouped all Malaysia hatchery populations into single clade. These tree topology structures suggest that the genetic variation within and between hatchery populations was low. These results were supported by the pairwise distance among hatchery populations, which return with no variations among hatchery populations. Furthermore, haplotype distribution also revealed that Malaysia hatchery populations were represented by single haplotype only. The situation was also observed from genetic diversity of *A. testudineus* Malaysia hatchery populations. There were no genetic variations observed among hatchery populations in this study.

According to Hidayat and Senanan (2010), there was low genetic variation within hatchery populations after a few generations of selection and domestication. This condition is due to improper hatchery and broodstock management. Some of the aquaculture practices that lead to loss of genetic is group spawning where the contribution of parent is practically unknown. Moreover, this practice makes it difficult for selection of the next generation, where the relationships and genetic lineages of produced offspring are unclear. This will lead to inbreeding which in turn contributes to loss of genetic variation within hatchery populations.

The reduction of genetic variations within hatchery populations has been recorded among other cultured fish species. Hatchery breeding program often uses small number of brood stock; this practice is primarily the cause of loss of genetic variations (Perez-Enriquez et al. 1999). When the number of brood stock was low, there are high chances of inbreeding, and the contribution of genetic variations is also low. In time, this practice may lead to low genetic variations within hatchery populations as demonstrated in this study. However, with a proper practice, mitigation on genetic loss can be considered. In red sea bream study (Perez-Enriquez et al. 1999), they have found that even though the genetic variations within hatchery populations were reduced, the genetic loss is not significant. This proofs that with proper management, the rate of genetic loss within hatchery populations can be reduced. This is very important for aquaculture industry to sustain for the future. If no mitigation measures were taken, the loss of genetic variations within hatchery populations may face severe effects.

There were practically no genetic differentiation and genetic diversity observed within hatchery populations of Malaysia *A. testudineus* in this study. The loss of genetic variation may be due to extensive use of the same population for multiple generations without introducing exotic brood stock to improve the genetic variations. In some hatchery of *Plecoglossus altivelis*, breeding this species has been

persistently using the same brood stock for multiple generations due to difficulty to collect mature brood stocks from natural populations (Iguchi et al. 1999). However, in this study it was not difficult to collect mature brood stocks from wild to improve genetic variations within hatchery populations. Demonstrated low genetic variations within hatchery populations might be due to the lack of information in genetic inventory of this species. Traditionally, to improve genetic gain in hatchery populations, brood stock from different hatcheries will be used in breeding program. Without proper information in genetic makeup of selected brood stocks, it may cause inbreeding or selection of same genetic populations which in turn decreases genetic variations within hatchery populations.

Genetic differentiations within hatchery populations typically reduced due to inappropriate hatchery management. The loss of genetic variations leads to detrimental effect on the performance of cultured fish species such as disease resistance, growth, and survival (Sekino et al. 2003). However, genetic gain leads to heterosis where offspring display exceed average performance of their parents (Lu et al. 2011).

In this study, there were high genetic variations among wild populations as well as within populations of *A. testudineus.* Due to high genetic variations, the data suggested that with proper hatchery management, selective breeding can be implemented to improve genetic variations within hatchery populations of *A. testudineus.* It is also possible to develop a better *A. testudineus* brood stock by crossbreed between distinct genetic lineages.

Phylogenetic analysis shows that all the other haplotypes were group in a same clade. However, the clade was consisted of two main branches. The Hap01 and Hap05 were grouped into top branch; meanwhile, Hap04 and Hap06 were grouped into lower branch. This might suggest that Hap02, Hap04, and Hap06 were from different genetic lineages. From phylogenetic analysis, Hap05 and Hap01 might share the same ancestor; thus, they are closely related. Therefore, Hap02, Hap04, and Hap06 can be used to enhance the hatchery populations. By introducing these haplotypes to the hatchery populations, it can improve genetic variations in the populations. The analysis of pairwise genetic distance among haplotypes suggested that there are a lot of potential candidates to improve genetic variations and genetic gain within *A. testudineus* hatchery populations. Pairwise genetic distance among haplotypes shows high segregation between Hap01 (hatchery population) and other haplotypes.

Selective breeding and genetic improvement within aquaculture industry have been introduced for Nile tilapia, salmonid species, and marine shrimp, *Penaeus vannamei* (Gjedrem et al. 2012). Genetic improvement in Atlantic salmon demonstrated improvement in gene variations within cultured salmon. In response to genetic gain, the cultured salmon demonstrated improvement in growth rate and feed conversion ratio (Gjedrem et al. 2012). This evident demonstrated that selective breeding and genetic improvement can benefit production within aquaculture industries. Basically, aquaculture industries gain benefit from selective breeding and genetic improvement program. Overall, implemented breeding program has produced improvement in growth rate and other commercial important trait.

In aquaculture industries, selective breeding has been used to gain selected commercial important trait. Without concerning genetic variations or genetic gain, improper selective breeding might contribute to loss of genetic variations within hatchery populations. However, studies on selective breeding evidently suggested that selection without genetic concern may not effect in improvement within cultured stock over the base population as observed in Nile tilapia. The lack of response to selection in this example can be attributed to genetic deterioration due to loss of genetic variations and inbreeding (Lind et al. 2012). Based on data obtained in this study, selective breeding with a proper guideline on genetic lineages of *A. testudineus* can be implemented. Genetic differentiation and phylogenetic information gathered in this study shed light for establishment of baseline populations and candidate for selection.

5 Conclusion

Analysis of cyt *b* mitochondrial DNA revealed that hatchery populations were representing by a single haplotype. This condition shows sign of genetic loss due to improper hatchery management and breeding practices. The selection and breeding were carried out without considering genetic background of selected candidates; if this practice persists it will lead to genetic deterioration due to inbreeding and loss of genetic variations. During this study, there is still no clear evidence of inbreeding depression or effect of genetic loss to the hatchery populations; however, a long-term exposure on inbreeding and loss of genetic will lead to devastating effect on performance and production of *A. testudineus* in Malaysia. If this situation persists longer, the loss of genetic variations will occur, and it will prevent selection for a better strain and it will produce progeny with developmental disorder. It is very important to develop hatchery populations with high genetic variations and to plan a proper breeding practice to avoid genetic loss in hatchery populations. In order to do so, this study has identified several sources in breeding practices that lead to genetic loss in hatchery populations and suggest mitigating measures to counter the effects. Existing hatchery populations are good candidates to be used as baseline populations.

Acknowledgment This research was conducted with joint funding from the Ministry of Education (MOE), under Fundamental Research Grant Scheme (FRGS) vote 59229 and also from IIUM funding (RIGS 16-100-0264).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410

Ambak MA, Isa MM, Zakaria MZ, Ghaffar MA (2010) Fishes of Malaysia. Penerbit Universiti Malaysia Terengganu

Bungas K, Arfiati D, Marsoedi, Halim H (2013) Effects of protein levels on the growth of climbing perch, *Anabas testudineus* Galam type, in Peat Water. Int Res J Biol Sci 2(4):55–58

Carr SM, Marshall HD (1991) Detection of intraspecific. DNA sequence variation in the mitochondrial cytochrome *b* gene of Atlantic Cod (*Gadus morhua*) by the polymerase chain reaction. Can J Fish Aquat Sci 48(1):48–52

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.5): an integrated software package for population genetics data analysis. Evol Bioinforma 1:47–50

Filgueira R, Grant J, Strand Ø, Asplinc L, Aurec J (2010) A simulation model of carrying capacity for mussel culture in a Norwegian fjord: role of induced upwelling. Aquaculture 308 (1–2):20–27

Gjedrem T, Robinson N, Rye M (2012) The importance of selective breeding in aquaculture to meet future demands for animal protein: a review. Aquaculture 350:117–129

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Hebert PD, Cywinska A, Ball SL (2003) Biological identifications through DNA barcodes. Proc R Soc Lond B Biol Sci 270(1512):313–321

Hidayat S, Senanan W (2010) PCR-RFLP analysis of mitochondrial DNA to differentiate populations of Climbing Perch (*Anabas testudineus*) in Thailand. Burapha Sci J 15(2):87–98

Iguchi K i, Watanabe K, Nishida M (1999) Reduced mitochondrial DNA variation in hatchery populations of ayu (*Plecoglossus altivelis*) cultured for multiple generations. Aquaculture 178 (3–4):235–243

Lind C, Ponzoni R, Nguyen N, Khaw H (2012) Selective breeding in fish and conservation of genetic resources for aquaculture. Reprod Domest Anim 47(4):255–263

Lu X, Wang H, Liu B, Xiang J (2011) An effective method for parentage determination of the clam (*Meretrix meretrix*) based on SSR and COI markers. Aquaculture 318(1):223–228

McCarthy C (1997) Chromas v1. 4 computer package. Griffith University, Australia

Perez-Enriquez R, Takagi M, Taniguchi N (1999) Genetic variability and pedigree tracing of a hatchery-reared stock of red sea bream (*Pagrus major*) used for stock enhancement, based on microsatellite DNA markers. Aquaculture 173(1–4):413–423

Rahman S, Monir MS, Khan MH (2013) Culture potentials of Climbing Perch, Thai Koi, *Anabas Testudineus* (Bloch) under different stocking densities in Northern Regions of Bangladesh. J Exp Biol Agric Sci 1(3):202–208

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19(18):2496–2497

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning, vol 2. Cold Spring Harbor Laboratory Press, New York, pp 14–19

Sekino M, Saitoh K, Yamada T, Kumagai A, Hara M, Yamashita Y (2003) Microsatellite-based pedigree tracing in a Japanese flounder *Paralichthys olivaceus* hatchery strain: implications for hatchery management related to stock enhancement program. *Aquaculture 221*(1-4):255–263

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24(8):1596–1599

Thompson JD (1991) Phenotypic plasticity as a component of evolutionary change. Trends Ecol Evol 6(8):246–249

Tingga RCT, Abdullah MT (2011) Phylogeny and phylogeography of *Aethalops* from Sundaland using mitochondrial 12S rRNA gene. Pertanika J Trop Agric Sci 35(3):485–510

Zalina I, Saad C, Christianus A, Harmin S (2012) Induced breeding and embryonic development of Climbing Perch (*Anabas testudineus*, Bloch). J Fish Aquat Sci 7(5):291–306

Molecular Identification of Reptiles from Tabuk Region of Saudi Arabia Through DNA Barcoding: A Case Study

Bishal Dhar, Mohua Chakraborty, N. Neelima Devi, Sorokhaibam Malvika, Madhurima Chakraborty, Subrata Trivedi, Abdulhadi A. Aloufi, and Sankar K. Ghosh

Abstract The deserts of Saudi Arabia provide an excellent habitat for reptiles. Although reptiles show significant vertebrate diversity, only few barcoding studies have been conducted on reptiles. In this case study, we collected different reptile species from the Tabuk region of Saudi Arabia and performed DNA barcoding in order to validate those species. We performed DNA sequencing for the COI region of 21 species belonging to the order Squamata. The BOLD Identification System (IDS) was used to establish species identity of the developed sequences. We searched both the private and published data in BOLD for available sequences through the "All Barcode Records" search engine. The neighbor-joining tree of all the species under this study was constructed, and the phylogenetic reconstruction was done using K2P distance model as per the standard protocol of DNA barcode. It was observed that *Chamaeleo chamaeleon* clusters with three *Diplometopon zarudnyi* sequences, of them two sequences have been generated in the lab and one sequence have been extracted from the database. *Eurylepis taeniolatus* also formed distinct branch in vicinity of three sequences of *Myrophis platyrhynchus.* This case study demonstrated the effectiveness of COI barcodes for reptile species from Saudi Arabia in discriminating species recognized through prior taxonomic work contributing to the growing library of DNA barcodes of animal species of the world. Some species groups with overlapping barcodes identified in this study were good candidates for further studies of phylogeography and speciation processes.

B. Dhar · M. Chakraborty · N. N. Devi · S. Malvika · M. Chakraborty
Department of Biotechnology, Assam University, Silchar, Assam, India

S. Trivedi
Department of Biology, University of Tabuk, Tabuk, Saudi Arabia

A. A. Aloufi
Department of Biology, University of Tabuk, Tabuk, Saudi Arabia

Department of Biology, Faculty of Science, Taibah University, Medinah, Saudi Arabia

S. K. Ghosh (✉)
Department of Biotechnology, Assam University, Silchar, Assam, India

University of Kalyani, Nadia, West Bengal, India

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_25

Further phylogenetic work on these species will reveal which of these highly divergent and geographically separated populations should be treated as belonging to the same species or sister species.

Keywords COI · DNA barcoding · BOLD-IDS · K2P · Reptiles · NJ

1 Reptiles: A Fundamental Component of Biodiversity

Reptiles are the group of vertebrates (they have a backbone) that include snakes, lizards, alligators, crocodiles, turtles, worm lizards, and caimans. The reptiles originated about 310–320 million years ago, in the swamps of the late Carboniferous period (Laurin and Reisz 1995). Reptiles are tetrapod vertebrates that either have four limbs or like snakes, which had descended from four-limbed ancestors. Reptiles, contrasting to amphibians, do not have aquatic larval stage. Most of them are oviparous (Sander 2012). Reptiles are important components of the food webs in most ecosystems. They fill up a critical role both as predator and prey. In some areas, they help control the numbers of serious agricultural pests by consuming rodent and insect pests. Reptiles have been hunted or traded, particularly as food (important protein source), traditional medicines, leathers as well as decorative (http://www.endangeredspeciesinternational.org/reptiles3.html). Modern-day reptiles (Squamata) comprising snakes and lizards are the most diverse order of reptiles with more than 9600 species (Sander 2012). All squamates shed their skin. Another characteristic that joins squamates is their jointed skulls and jaws, which enable them to open very wide and consume prey much larger than themselves. There are more than 4500 species of lizards, about 2900 species of snakes, and 130 species of worm lizards (http://study.com/academy/lesson/reptiles-features-types-importance.html).

Saudi Arabia Saudi Arabia occupies most of the Arabian Peninsula (80%), with the Red Sea and the Gulf of Aqaba to the west and the Persian Gulf to the east. Saudi Arabia contains the world's largest continuous sand desert, the Rub Al-Khali, or Empty Quarter. It has a land area of 829,995 sq mile (2,149,690 sq km) (the area between long. 44°30′−56°30′E and lat. 16°30′−23°00′N) (http://www.factmonster.com/country/saudi-arabia.html). The Arabian Desert has a subtropical, hot desert climate, close to the climate of the Sahara Desert, the world's largest hot desert located in North Africa. In fact, the Arabian Desert is an extension of the Sahara Desert over the Arabian Peninsula. The climate is mainly hot and dry with plenty of sunshine throughout the year. The temperatures oscillate between very high heat and seasonal nighttime freezes. The desert of Saudi Arabia provides an excellent place for reptiles from the savage extremes of climate, because even a few inches of sand offer excellent insulation against heat and cold (http://www.saudiaramcoworld.com/issue/196805/the.toadhead.from.najad.and.other.reptiles.htm) (Fig. 1).

DNA Barcoding and Species Identification The ability to accurately identify and describe species underpins all biological research, yet the traditional morphological-

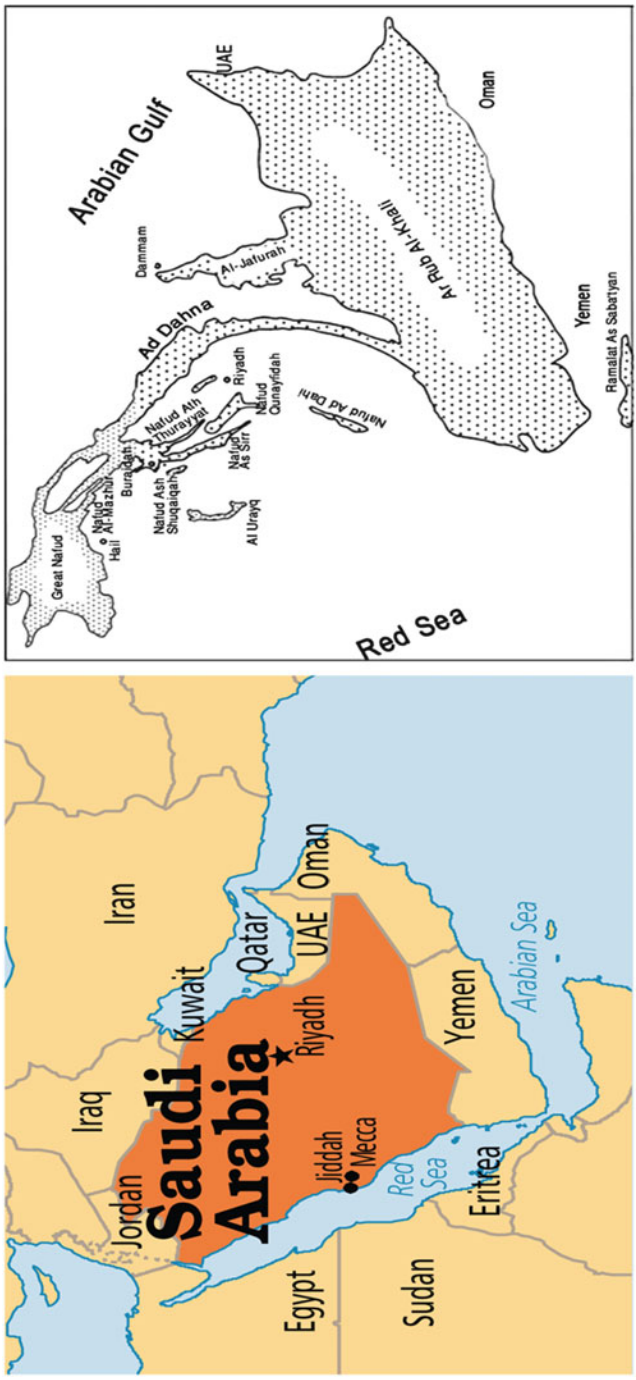


Fig. 1 Study site (Saudi Arabia) (http://www.operationworld.org/saud)

based taxonomic approaches have only managed to describe 1.2–1.5 million species over the past 250 years (Mora et al. 2011; Chapple and Ritchie 2013), a mere 10% of the Earth's predicted eukaryotic diversity (Mora et al. 2011). It is estimated that persisting with such time, consuming and cumbersome approaches would not result in a comprehensive inventory of the world's biodiversity for at least ~1000 years (Packer et al. 2009; Chapple and Ritchie 2013) and perhaps much longer given the sharp decline in the number of specialist taxonomists and funding for taxonomic research (Rodman and Cody 2003; Wheeler et al. 2004). The DNA barcoding approach was introduced in 2003 by Paul Hebert and colleagues (Hebert et al. 2003), as a way to overcome the existing taxonomic "impediment" or "bottleneck" (Hebert et al. 2003). DNA barcoding has become a promising tool for the rapid and accurate identification of various species and managing species diversity (Hebert et al. 2003; Dawnay et al. 2007; Trivedi et al. 2014, 2015, 2016a–c; Dhar et al. 2016a, b). It promised to revolutionize the identification of existing species and speed the discovery of new species. DNA barcoding can be helpful in species diagnosis because sequence divergences are ordinarily much lower among individuals of a species than between closely related species (Hebert et al. 2003). The distinction between intra- and inter-specific divergences, termed the "barcoding gap" (Meyer and Paulay 2005), enables unknown sequences to be assigned to an existing species or flagged as a suspected new species. DNA barcoding employs sequence diversity in short, standardized gene regions to aid species identification and discovery in large assemblages of life. A 648-bp region of the cytochrome c oxidase I (*COI*) gene forms the primary barcode sequence for members of the animal kingdom (Hebert et al. 2003; Savolainen et al. 2005). A significant advantage of the DNA barcoding approach is that it works in situations that would confound many morphological approaches: specimen fragments (Armstrong and Ball 2005; Chapple and Ritchie 2013), species with multiple life stages (Hebert et al. 2004), and sexual dimorphism or conserved, variable or plastic morphology (Smith et al. 2006, 2007; Burns et al. 2008). DNA barcoding is not only a powerful tool for species identification but also can play a vital role in wildlife forensics and conservation genetics (Wolinsky 2012). This technique is widely used for the authentication and safety assessment of food (Saggu et al. 2016). The occurrence of cryptic species is relatively common in nature. Cryptic species are those species that are morphologically similar but genetically distinct. DNA barcoding can be a very effective tool in assessment of these cryptic species (Hebert et al. 2004). DNA barcoding can also be very effective for molecular phylogenetic studies (Ajmal Ali et al. 2014; Trivedi et al. 2016a). Several bioinformatics tools are available nowadays for efficient DNA barcoding (Mahadani et al. 2016). DNA barcoding allows a day to be envisioned when every curious mind, from professional biologists to school children, will have easy access to the names and biological attributes of any species on the planet.

2 DNA Barcoding of Reptiles

DNA barcoding like other species is also used to identify reptile species and assess the taxonomic position by a number of studies. Few years back by a group of workers, the classification of *Terrapene* spp. (American box turtles) was reassessed by obtaining DNA sequence data from a broad geographic range and from all 4 recognized species and 11 subspecies within the genus (Martin et al. 2013). The mtDNA *COI* gene was sequenced to calculate the genetic divergence for each *Terrapene taxon.* Among the studies species *Terrapene carolina mexicana* and *T. c. yucatana* formed a monophyletic clade with *T. c. triunguis*, and this clade was paraphyletic to the rest of *T. carolina. T. ornata ornata* and *T. o. luteola* lacked distinction phylogenetically, and *Terrapene nelsoni* was confirmed to be the sister taxon of *T. ornata*. The DNA barcoding results indicated that all taxa were different species (>2% sequence divergence) except for *T. c. triunguis–T. c. mexicana* and *T. o. ornata–T. o. luteola*. The results suggest that *T. c. triunguis* should be elevated to species status (*Terrapene mexicana*), and mexicana and yucatana should be included in this group as subspecies. Vargas et al. (2009) generated *COI* sequences for the five species of sea turtles that occur in Brazil. These presented widely divergent haplotypes. All observed values were on the same range as those already described for other animal groups: the overall mean distance was 8.2%, the mean distance between families (Dermochelyidae and Cheloniidae) 11.7%, the mean intraspecific divergence 0.34%, and the mean distance within Cheloniidae 6.4%, this being 19-fold higher than the mean divergence observed within species. The finding of characteristic species-specific *COI* sequences offers the prospect of identifying marine turtle species by using DNA barcode methodology as an auxiliary tool for taxonomy. Nagy et al. (2012) carried out the first large-scale DNA barcoding of reptiles (including Squamata and Testudines) with 468 specimens from biodiversity hotspot of Madagascar. In this study 41–48 new (undescribed) species were identified thereby indicating the utility of DNA barcoding in biodiversity assessment. The study also revealed that the average interspecific genetic distance within families was 13.4% in Boidae and 29.8% in Gekkonidae. Hawlitschek et al. (2013) revealed the results of a DNA barcoding study of squamates of the Comoros archipelago. The barcoding dataset included 27 of the 29 currently recognized squamate species of the Comoros, including 17 of the 18 endemic species. Some species considered endemic to the Comoros according to current taxonomy were found to cluster with non-Comoran lineages, probably due to poorly resolved taxonomy. All other species for which more than one barcode was obtained corresponded to distinct clusters useful for species identification by barcoding. Two cryptic species were identified using the DNA barcoding approach. The study inferred that in the case of the squamates of the Comoros Islands, DNA barcoding proved to be a very useful and efficient way of detecting isolated populations and promising starting points for subsequent research. Nguyen et al. (2014) used DNA barcoding using *COI* (550 bp) to reexamine the levels of genetic divergence and taxonomic status of 21 described species of Vietnamese bent-toed geckos. Tree-

based analyses resolved all sampled species and identified potential undescribed taxa. Further, their analyses discovered two potentially new species from Vietnam, two from Laos, and one from China. Huang et al. (2015) discussed the feasibility of DNA barcoding method for identification of commonly used medicinal snakes. The phylogenetic trees using neighbor joining were constructed. As calculated by Kimura 2-parameter model, the mean intraspecies genetic distance of *Trimeresurus albolabris, Ptyas dhumnades*, and *Lycodon rufozonatus* was greater than 2%. Further phylogenetic relationship results suggested that identification of one sample of *T. albolabris* was erroneous. The identification of some samples of *P. dhumnades* was also not correct, namely, originally *P. korros* was identified as *P. dhumnades*. Therefore, DNA barcoding for identification of medicinal snakes is feasible and greatly complements the morphological classification method. Recently, a review article has been published on DNA barcoding of different groups of reptiles (Trivedi et al. 2016d).

3 Identification of Reptiles of Tabuk Through DNA Barcoding: BLAST Result Analysis

Twenty-one reptile sequences from the order Squamata have been collected from Tabuk and sequenced. The BLAST search results of these sequences have been detailed in Table 1. A neighbor-joining (NJ) tree has been constructed using the developed sequences along with the downloaded BLAST hits of individual sequences. Only those BLAST hits have been considered which have the highest scores, and E_value is close to 0. Among them, only eight sequences have conspecific sequences available in the database. Remaining sequences showed match with the closest available relative in the database like congeneric or confamilial species. In some rare cases, in the absence of true phylogenetic relative in the database, the closest hit showed random matches with species belonging to completely different taxa, like Aves and Anguilliformes. However these cases were associated with high E value which makes the hit false positive. As in the case of *Acanthodactylus opheodurus,* in the absence of conspecific sequence, BLAST generated hit with 98% query coverage and 82% similarity with conger sequences which belong to the phylum Aves. The E value of the match was however high with 1.00E-125 that showed random match. The taxonomic details of BLAST hits have been given in the Table 2.

4 Species Identification Using BOLD

The BOLD Identification System (IDS) was used to establish species identity of the developed sequences. This identification system for COI accepts sequences from the 5′ region of the mitochondrial *Cytochrome c oxidase subunit I* gene and returns a

Table 1 Similarity match with GenBank sequences using nucleotide BLAST

Sample code	Reptiles vouchered	Species match in BLAST	E value	Identity	Accession
001(F)	*Chamaeleo chamaeleon*	*Diplometopon zarudnyi* voucher MVZ 234273	0	99%	AY605474.1
2R(F)	*Chalcides ocellatus*	*Sceloporus virgatus* voucher AMNH Herpetology 137700	2.00E-133	82%	KU985944.1
		Sceloporus virgatus voucher AMNH Herpetology 137699	2.00E-133	82%	KU985908.1
		Hydrobates pelagicus voucher NHMO-BC33	1.00E-130	82%	GU571435.1
		Hydrobates pelagicus voucher NHMO-BC32	1.00E-130	82%	GU571434.1
3R(F)	*Scincus mitranus*	*Oligosoma maccanni* isolate OMA7	1.00E-125	82%	KC349736.1
		Oligosoma maccanni isolate OMA2	1.00E-125	82%	KC349722.1
		Oligosoma maccanni isolate OMA15	1.00E-125	82%	KC349720.1
5R(F)	*Eurylepis taeniolatus*	*Myrophis platyrhynchus* voucher MFL356	2.00E-133	82%	GU224964.1
		Myrophis platyrhynchus voucher MFL354	2.00E-133	82%	GU224963.1
		Myrophis platyrhynchus voucher MFL353	2.00E-133	82%	GU224956.1
7(F)	*Stellagama stellio*	*Stellagama stellio* voucher ZMMU R-11324	0	92%	KF691700.1
8(F)	*Stellagama stellio*	*Stellagama stellio* voucher ZMMU R-11324	0	91%	KF691700.1
009(F)	*Pseudotrapelus aqabensis*	*Pseudotrapelus aqabensis* isolate C-5-33	0	100%	KP994947.1
		Pseudotrapelus dhofarensis isolate C-4-242743	0	91%	KP994946.1
		Pseudotrapelus jensvindumi isolate C-7-236932	0	90%	KP994949.1
		Pseudotrapelus jensvindumi voucher CAS:225340	0	90%	KP979760.1
		Pseudotrapelus dhofarensis voucher ZISP:26351	0	90%	KP979759.1
10(F)	*Pseudotrapelus aqabensis*	*Pseudotrapelus aqabensis* isolate C-5-33	0	99%	KP994947.1
		Pseudotrapelus dhofarensis isolate C-4-242743	0	91%	KP994946.1
		Pseudotrapelus jensvindumi isolate C-7-236932	0	90%	KP994949.1
		Pseudotrapelus jensvindumi voucher CAS:225340	0	90%	KP979760.1

(continued)

Table 1 (continued)

Sample code	Reptiles vouchered	Species match in BLAST	E value	Identity	Accession
12(F)	*Diplometopon zarudnyi*	*Diplometopon zarudnyi* voucher MVZ 234273	0	99%	AY605474.1
13(F)	*Rhagerhis moilensis*	*Mimophis mahfalensis* voucher REPT_M12473	2.00E-167	86%	JQ909478.1
16(F)	*Cerastes gasperettii*	*Cerastes cerastes*	0	89%	EU852311.1
19(F)	*Cyrtopodion scabrum*	*Auriparus flaviceps* voucher FMNH 394359	1.00E-111	80%	DQ432755.1
		Hemidactylus pumilio voucher IBES5021	1.00E-110	80%	KU567474.1
21(F)	*Stenodactylus doriae*	*Cephalopholis cyanostigma* voucher UG0456	1.00E-111	80%	KP194176.1
22(F)	Stenodactylus doriae	*Cyanopica cyanus* isolate: YIO318-10	2.00E-118	80%	AB843453.1
25(F)	*Mesalina brevirostris*	*Monasa morphoeus* voucher LGEMA-3306	1.00E-125	82%	JN801821.1
		Monasa morphoeus voucher LGEMA-9860	1.00E-120	81%	JN801823.1
26R(F)	*Acanthodactylus opheodurus*	*Conger conger* voucher CSFOM-031	1.00E-125	82%	KJ709504.1
		Conger conger voucher Re 2 Hg 190506 E	1.00E-125	82%	JN231238.1
		Conger conger voucher FCFOPB064-17	1.00E-125	82%	JQ775006.1
27(F)	*Phoenicolacerta kulzeri khazaliensis*	*Phoenicolacerta kulzeri*	0	89%	FJ460596.1
29(F)	*Acanthodactylus opheodurus*	*Monasa morphoeus* voucher LGEMA-3428	2.00E-129	82%	JN801822.1
30(F)	*Hemidactylus flaviviridis*	*Hemidactylus homoeolepis* voucher CN1034	1.00E-160	85%	KU567377.1
060(F)	*Diplometopon zarudnyi*	*Diplometopon zarudnyi* voucher MVZ 234273	0	99%	AY605474.1
063(F)	*Stellagama stellio*	*Cerastes cerastes*	0	89%	EU852311.1

The result showed the closest match with the available database sequence. The similarity of the sequences was expressed in terms of percentage of identity with E value

species-level identification when one is possible. We searched both the private and published data in BOLD for available sequences through the "All Barcode Records" search engine. The search returns every COI barcode record on BOLD with a minimum sequence length of 500 bp including unvalidated library and records without species-level identification. This also includes many species represented by only one or two specimens as well as all species with interim taxonomy. Further

Table 2 Taxonomic details of the BLAST hit results in NCBI

BLAST hits	Taxonomy
AY605474.1|*Diplometopon zarudnyi*	Chordata; Reptilia; Squamata; Trogonophidae; Diplometopon
DQ432755.1|*Auriparus flaviceps*	Chordata; Aves; Passeriformes; Remizidae; Auriparus
EU852311.1|*Cerastes cerastes*	Squamata; Viperidae
FJ460596.1|*Phoenicolacerta kulzeri*	Squamata; Lacertidae; Phoenicolacerta
GU571434.1|*Hydrobates pelagicus*	Chordata; Aves; Procellariiformes; Hydrobatidae; Hydrobates
GU571435.1|*Hydrobates pelagicus*	Chordata; Aves; Procellariiformes; Hydrobatidae; Hydrobates
GU224956.1|*Myrophis platyrhynchus*	Chordata; Actinopterygii; Anguilliformes; Ophichthidae; Myrophinae; Myrophis
GU224963.1|*Myrophis platyrhynchus*	Chordata; Actinopterygii; Anguilliformes; Ophichthidae; Myrophinae; Myrophis
GU224964.1|*Myrophis platyrhynchus*	Chordata; Actinopterygii; Anguilliformes; Ophichthidae; Myrophinae; Myrophis
JN231238.1|*Conger conger*	Chordata; Actinopterygii; Anguilliformes; Congridae; Congrinae; Conger
JN801821.1|*Monasa morphoeus*	Chordata; Aves; Galbuliformes; Bucconidae; Monasa
JN801822.1|*Monasa morphoeus*	Chordata; Aves; Galbuliformes; Bucconidae; Monasa
JN801823.1|*Monasa morphoeus*	Chordata; Aves; Galbuliformes; Bucconidae; Monasa
JQ909478.1|*Mimophis mahfalensis*	Chordata; Reptilia; Squamata; Lamprophiidae; Psammophiinae; Mimophis
JQ775006.1|*Conger conger*	Chordata; Reptilia; Squamata; Lamprophiidae; Psammophiinae; Mimophis
KC349720.1|*Oligosoma maccanni*	Scincidae
KC349722.1|*Oligosoma maccanni*	Scincidae
KC349736.1|*Oligosoma maccanni*	Scincidae
KF691700.1|*Stellagama stellio*	Chordata; Reptilia; Squamata; Agamidae; Agaminae; Stellagama
AB843453.1|*Cyanopica cyanus*	Chordata; Aves; Passeriformes; Corvidae; Cyanopica
KJ709504.1|*Conger conger*	Chordata; Aves; Passeriformes; Corvidae; Cyanopica
KP979759.1|*Pseudotrapelus dhofarensis*	Chordata; Reptilia; Squamata; Agamidae; Agaminae; Pseudotrapelus
KP979760.1| *Pseudotrapelus_jensvindumi*	Chordata; Reptilia; Squamata; Agamidae; Agaminae; Pseudotrapelus
KP994946.1|*Pseudotrapelus dhofarensis*	Chordata; Reptilia; Squamata; Agamidae; Agaminae; Pseudotrapelus

(continued)

Table 2 (continued)

BLAST hits	Taxonomy
KP994947.1|*Pseudotrapelus aqabensis*	Chordata; Reptilia; Squamata; Agamidae; Agaminae; Pseudotrapelus
KP994949.1|*Pseudotrapelus jensvindumi*	Chordata; Reptilia; Squamata; Agamidae; Agaminae; Pseudotrapelus
KP194176.1| *Cephalopholis_cyanostigma*	Chordata; Actinopterygii; Perciformes; Serranidae; Epinephelinae; Cephalopholis
KU567377.1|*Hemidactylus homoeolepis*	Chordata; Reptilia; Squamata; Gekkonidae; Hemidactylus
KU567474.1|*Hemidactylus pumilio*	Chordata; Reptilia; Squamata; Gekkonidae; Hemidactylus
KU985908.1| *Sceloporus_virgatus*	Chordata; Reptilia; Squamata; Phrynosomatidae; Sceloporinae; Sceloporus

the "Species Level Barcode Records" was used to extract a list of the nearest matches and that provided a probability of placement to a taxon.

Among the 21 COI barcode sequences developed in the lab, species status for only 5 sequences could be confirmed using BOLD Identification System. For most of the remaining sequences, conspecific sequences were not available in the BOLD database. Table 3 shows detailed description of similarity match of the sequences using BOLD Identification System. Top five matches of the sequences using the "All Barcode Records" search were displayed for each of the sequence. In case of 001 (F) *Chamaeleo chamaeleon*, 15 COI sequences were available in the BOLD database. However, the top five similarity match did not show close identity with any of these sequences. Instead the sequence showed 99.81% similarity with *Diplometopon zarudnyi*, and IDS identified the sequence as *Diplometopon zarudnyi*. Such incongruency in the similarity may be because of the presence of hybrid sequences or mislabelled sequence. Conspecific sequences for 2R(F) *Chalcides ocellatus* were not available in the BOLD database. 3R(F) *Scincus mitranus* showed 95.4% similarity with congeneric sequence *Scinus scinus* available in private database. Three sequences of *Eurylepis taeniolatus* were found in early release section; however they showed an average of 87% match with the 5R(F) *Eurylepis taeniolatus.* Four sequences of *Stellagama stellio* were present in the database. They showed 88–96% similarity with 7(F) *Stellagama stellio*, and IDS did not identify species status of the sequence. However, 8(F) *Stellagama stellio* was identified up to species level as it showed 98.5% similarity with database conspecific sequence. Developed sequences of *Pseudotrapelus aqabensis* 9(F) and 10(F) showed 99% similarity with database sequences and were identified correctly by IDS. 12(F) *Diplometopon zarudnyi* showed 99% similarity with database sequence and was identified correctly up to species level. *Rhagerhis moilensis* and *Mesalina brevirostris* did not have any conspecific sequences available in the database. However, *Mesalina brevirostris* showed 98% similarity with *Acanthodactylus boskianus* and hence was identified as same species. *Cyrtopodion scabrum* has a conspecific sequence available in the database, but IDS did not show significant similarity with this sequence.

Table 3 Species identification using BOLD-IDS (Barcode of Life Data System-Identification system) search engine

Voucher ID	Vouchered specimen	Top hit (similarity)	Status	Species identification
001(F)	*Chamaeleo chamaeleon*	*Diplometopon zarudnyi* (99.81)	Private	No
2R(F)	*Chalcides ocellatus*	No match		No
3R(F)	*Scincus mitranus*	No match		No
5R(F)	*Eurylepis taeniolatus*	No match		No
7(F)	*Stellagama stellio*	No match		No
8(F)	*Stellagama stellio*	*Stellagama stellio* (98.51)	Early release	Species identified
009(F)	*Pseudotrapelus aqabensis*	*Pseudotrapelus aqabensis* (98.9)	Published	Species identified
10(F)	*Pseudotrapelus aqabensis*	*Pseudotrapelus aqabensis* (99.45)	Published	Species identified
12(F)	*Diplometopon zarudnyi*	*Diplometopon zarudnyi* (99.82)	Private	Species identified
13(F)	*Rhagerhis moilensis*	No match		No
16(F)	*Cerastes gasperettii*	No match		No
19(F)	*Cyrtopodion scabrum*	No match		No
21(F)	*Stenodactylus doriae*	No match		No
22(F)	*Stenodactylus doriae*	No match		No
25(F)	*Mesalina brevirostris*	*Acanthodactylus boskianus* (98.38)	Early release	No
26R(F)	*Acanthodactylus opheodurus*	No match		No
27(F)	*Phoenicolacerta kulzeri khazaliensis*	No match		No
29(F)	*Acanthodactylus opheodurus*	*Acanthodactylus boskianus* (99.46)	Early release	Genus identified
30(F)	*Hemidactylus flaviviridis*	No match		No
060(F)	*Diplometopon zarudnyi*	*Diplometopon zarudnyi* (99.48)	Private	Species identified
063(F)	*Stellagama stellio*	No match		No

The developed sequences of the specimen were checked for similarity match in the Public Record Barcode Database of BOLD-IDS for comprehensive species identification

21(F) *Stenodactylus doriae* showed 81–89% similarity with the available database sequences, while 22(F) *Stenodactylus doriae* showed 91% similarity with the sequences. 63(F) *Stellagama stellio* did not show match with any of the available database sequences.

5 Neighbor-Joining Clustering

The neighbor-joining tree of all the species under this study is constructed as shown in Fig. 2. The phylogenetic reconstruction was done using K2P distance model as per the standard protocol of DNA barcode. As per the observation in this case, 001F *Chamaeleo chamaeleon* clusters have three *Diplometopon zarudnyi* sequences, of them two sequences [12(F) and 60(F)] have been generated in the lab and one sequence, AY605474, has been extracted from the database. Such clustering could be possible because of either the presence of mislabelled or misidentified sequence or there could be the possibility of species introgression. 2R(F) *Chalcides ocellatus* clusters separately as no conspecific sequence is available in database. However, they align close to (KU985908, KU985944) *Sceloporus virgatus* belonging to the same order Squamata but different family Phrynosomatidae. 3RF_*Scincus mitranus* clusters separately but closes to three confamilial database sequences of *Oligosoma maccanni* (KC349720, KC349736, KC349722). *Eurylepis taeniolatus* also forms distinct branch in vicinity of three sequences of *Myrophis platyrhynchus* (GU224956, GU224963-64), which are Anguilliformes. 7R and 8R *Stellagama stellio* cluster together along with another database sequence (KF691700) of the same species. However 63R *Stellagama stellio* clusters separately and closes to 16(F) *Cerastes gasperettii*. 009F and10R *Pseudotrapelus aqabensis* cluster together with conspecific sequence KP994947 from database. Moreover, four database sequences (KP979760, KP994949, KP979759, KP994946) from three congeneric species of Pseudotrapelus cluster distinctly under the same node. As conspecific sequences are not present in the database, 13 (F) *Rhagerhis moilensis* shows closest hit with *Mimophis mahfalensis* which belongs to the same family. In the NJ tree as well, the two sequences form close cluster distinct from other families. 19(F) *Cyrtopodion scabrum* forms subcluster with three sequences of Hemidactylus genus where sequences (KU567377, KU567474) of two species were extracted from the database and one sequence, 30(F) *Hemidactylus flaviviridis,* was developed in the lab. Both of these genera belong to the same family Gekkonidae. 21 (F) and 22(F) *Stenodactylus doriae* cluster together along with other sequences of Gekkonidae family. Species of Lacertidae family, 25(F) *Mesalina brevirostris* and 26R(F), 29(F) *Acanthodactylus opheodurus*, form distinct cluster. However 27 (F) *Phoenicolacerta kulzeri khazaliensis* ssp. forms separate cluster along with a conspecific database sequence FJ460596.

This case study demonstrated the effectiveness of COI barcodes for reptile species from Saudi Arabia in discriminating species recognized through prior taxonomic work contributing to the growing library of DNA barcodes of animal species of the world. The study showed that the partial COI gene enables accurate animal species identification where adequate reference sequence data exists. Some species groups with overlapping barcodes identified in this study were good candidates for further studies of phylogeography and speciation processes. Further phylogenetic work on these species will reveal which of these highly divergent and geographically separated populations should be treated as belonging to the same species or sister species.

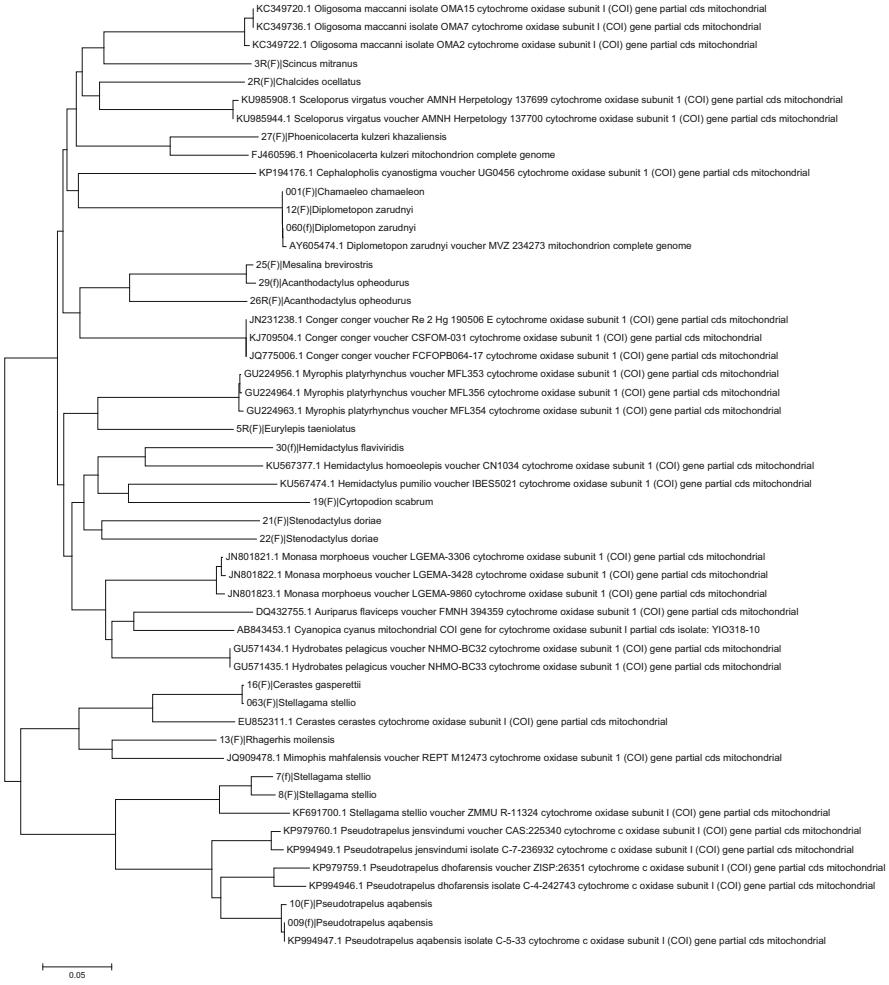

KC349720.1 Oligosoma maccanni isolate OMA15 cytochrome oxidase subunit I (COI) gene partial cds mitochondrial
KC349736.1 Oligosoma maccanni isolate OMA7 cytochrome oxidase subunit I (COI) gene partial cds mitochondrial
KC349722.1 Oligosoma maccanni isolate OMA2 cytochrome oxidase subunit I (COI) gene partial cds mitochondrial
3R(F)|Scincus mitranus
2R(F)|Chalcides ocellatus
KU985908.1 Sceloporus virgatus voucher AMNH Herpetology 137699 cytochrome oxidase subunit 1 (COI) gene partial cds mitochondrial
KU985944.1 Sceloporus virgatus voucher AMNH Herpetology 137700 cytochrome oxidase subunit 1 (COI) gene partial cds mitochondrial
27(F)|Phoenicolacerta kulzeri khazaliensis
FJ460596.1 Phoenicolacerta kulzeri mitochondrion complete genome
KP194176.1 Cephalopholis cyanostigma voucher UG0456 cytochrome oxidase subunit 1 (COI) gene partial cds mitochondrial
001(F)|Chamaeleo chamaeleon
12(F)|Diplometopon zarudnyi
060(f)|Diplometopon zarudnyi
AY605474.1 Diplometopon zarudnyi voucher MVZ 234273 mitochondrion complete genome
25(F)|Mesalina brevirostris
29(f)|Acanthodactylus opheodurus
26R(F)|Acanthodactylus opheodurus
JN231238.1 Conger conger voucher Re 2 Hg 190506 E cytochrome oxidase subunit 1 (COI) gene partial cds mitochondrial
KJ709504.1 Conger conger voucher CSFOM-031 cytochrome oxidase subunit 1 (COI) gene partial cds mitochondrial
JQ775006.1 Conger conger voucher FCFOPB064-17 cytochrome oxidase subunit 1 (COI) gene partial cds mitochondrial
GU224956.1 Myrophis platyrhynchus voucher MFL353 cytochrome oxidase subunit 1 (COI) gene partial cds mitochondrial
GU224964.1 Myrophis platyrhynchus voucher MFL356 cytochrome oxidase subunit 1 (COI) gene partial cds mitochondrial
GU224963.1 Myrophis platyrhynchus voucher MFL354 cytochrome oxidase subunit 1 (COI) gene partial cds mitochondrial
5R(F)|Eurylepis taeniolatus
30(f)|Hemidactylus flaviviridis
KU567377.1 Hemidactylus homoeolepis voucher CN1034 cytochrome oxidase subunit 1 (COI) gene partial cds mitochondrial
KU567474.1 Hemidactylus pumilio voucher IBES5021 cytochrome oxidase subunit 1 (COI) gene partial cds mitochondrial
19(F)|Cyrtopodion scabrum
21(F)|Stenodactylus doriae
22(F)|Stenodactylus doriae
JN801821.1 Monasa morphoeus voucher LGEMA-3306 cytochrome oxidase subunit 1 (COI) gene partial cds mitochondrial
JN801822.1 Monasa morphoeus voucher LGEMA-3428 cytochrome oxidase subunit 1 (COI) gene partial cds mitochondrial
JN801823.1 Monasa morphoeus voucher LGEMA-9860 cytochrome oxidase subunit 1 (COI) gene partial cds mitochondrial
DQ432755.1 Auriparus flaviceps voucher FMNH 394359 cytochrome oxidase subunit 1 (COI) gene partial cds mitochondrial
AB843453.1 Cyanopica cyanus mitochondrial COI gene for cytochrome oxidase subunit I partial cds isolate: YIO318-10
GU571434.1 Hydrobates pelagicus voucher NHMO-BC32 cytochrome oxidase subunit 1 (COI) gene partial cds mitochondrial
GU571435.1 Hydrobates pelagicus voucher NHMO-BC33 cytochrome oxidase subunit 1 (COI) gene partial cds mitochondrial
16(F)|Cerastes gasperettii
063(F)|Stellagama stellio
EU852311.1 Cerastes cerastes cytochrome oxidase subunit I (COI) gene partial cds mitochondrial
13(F)|Rhagerhis moilensis
JQ909478.1 Mimophis mahfalensis voucher REPT M12473 cytochrome oxidase subunit 1 (COI) gene partial cds mitochondrial
7(f)|Stellagama stellio
8(F)|Stellagama stellio
KF691700.1 Stellagama stellio voucher ZMMU R-11324 cytochrome oxidase subunit I (COI) gene partial cds mitochondrial
KP979760.1 Pseudotrapelus jensvindumi voucher CAS:225340 cytochrome c oxidase subunit I (COI) gene partial cds mitochondrial
KP994949.1 Pseudotrapelus jensvindumi isolate C-7-236932 cytochrome c oxidase subunit I (COI) gene partial cds mitochondrial
KP979759.1 Pseudotrapelus dhofarensis voucher ZISP:26351 cytochrome c oxidase subunit I (COI) gene partial cds mitochondrial
KP994946.1 Pseudotrapelus dhofarensis isolate C-4-242743 cytochrome c oxidase subunit I (COI) gene partial cds mitochondrial
10(F)|Pseudotrapelus aqabensis
009(f)|Pseudotrapelus aqabensis
KP994947.1 Pseudotrapelus aqabensis isolate C-5-33 cytochrome c oxidase subunit I (COI) gene partial cds mitochondrial
0.05

Fig. 2 Neighbor-joining tree of COI sequences of all the reptile species from Tabuk along with the other database sequences as study replicates

Acknowledgment We acknowledge the Deanship of Scientific Research, University of Tabuk, Saudi Arabia, for the support provided through the Project No. S-1437-0067.

References

Ajmal Ali M, Gyulai G et al (2014) The changing epitome of species identification – DNA barcoding. Saudi J Biol Sci 21(3):204–231

Armstrong KF, Ball SL (2005) DNA barcodes for biosecurity: invasive species identification. Philos Trans R Soc Lond B Biol Sci 360(1462):1813–1823

Burns JM, Janzen DH et al (2008) DNA barcodes and cryptic species of skipper butterflies in the genus Perichares in Area de Conservacion Guanacaste, Costa Rica. Proc Natl Acad Sci U S A 105(17):6350–6355

Chapple DG, Ritchie PA (2013) A retrospective approach to testing the DNA barcoding method. PLoS One 8(11):e77882

Dawnay N, Ogden R et al (2007) Validation of the barcoding gene COI for use in forensic genetic species identification. Forensic Sci Int 173(1):1–6

Dhar B, Ghose A, Kundu S, Malvika S, Devi NN, Choudhury A, Gorai S, Trivedi S, Ghosh SK (2016a) DNA barcoding: molecular positioning of living fossils (horseshoe crab). In: Trivedi S, Ansari AA, Ghosh SK, Rehman H (eds) DNA barcoding in marine perspectives. pp 181–199. https://doi.org/10.1007/978-3-319-41840-7_12

Dhar B, Ghose A, Kundu S, Choudhury A, Gorai S, Trivedi S, Nath J, Ghosh SK (2016b) DNA barcoding of primitive species-nemertine from Sundarbans marine bio-resourcee In: Trivedi S, Ansari AA, Ghosh SK, Rehman H (eds) DNA barcoding in marine perspectives. pp 157–168. https://doi.org/10.1007/978-3-319-41840-7_10

Hawlitschek O, Nagy ZT et al (2013) Reliable DNA barcoding performance proved for species and island populations of comoran squamate reptiles. PLoS One 8(9):e73368

Hebert PD, Cywinska A et al (2003) Biological identifications through DNA barcodes. Proc Biol Sci 270(1512):313–321

Hebert PD, Penton EH et al (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. Proc Natl Acad Sci U S A 101 (41):14812–14817

Huang Y, Zhang YY et al (2015) DNA barcoding and its utility in commonly-used medicinal snakes. Zhongguo Zhong Yao Za Zhi 40(5):868–874

Laurin M, Reisz RR (1995) A reevaluation of early amniote phylogeny. Zool J Linnean Soc 113 (2):165–223

Mahadani P, Trivedi S, Rehman H, Saggu S (2016) Bioinformatics tools in marine DNA barcoding. In: Trivedi S, Ansari AA, Ghosh SK, Rehman H (eds) DNA barcoding in marine perspectives. pp 71–85. https://doi.org/10.1007/978-3-319-41840-7_5

Martin BT, Bernstein NP et al (2013) Sequence-based molecular phylogenetics and phylogeography of the American box turtles (Terrapene spp.) with support from DNA barcoding. Mol Phylogenet Evol 68(1):119–134

Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. PLoS Biol 3(12):e422

Mora C, Tittensor DP et al (2011) How many species are there on Earth and in the Ocean? PLoS Biol 9(8):e1001127

Nagy ZT, Sonet G et al (2012) First large-scale DNA barcoding assessment of reptiles in the biodiversity hotspot of Madagascar, based on newly designed COI primers. PLoS One 7(3): e34506

Nguyen SN, Yang YX et al (2014) DNA barcoding of Vietnamese bent-toed geckos (Squamata: Gekkonidae: Cyrtodactylus) and the description of a new species. Zootaxa 3784:48–66

Packer L, Gibbs J et al (2009) DNA barcoding and the mediocrity of morphology. Mol Ecol Resour 9(Suppl S1):42–50

Rodman JE, Cody JH (2003) The taxonomic impediment overcome: NSF's Partnerships for Enhancing Expertise in Taxonomy (PEET) as a model. Syst Biol 52(3):428–435

Saggu S, Rehman H, Trivedi S, Aziz AT, Eldiasty JG (2016) Safety assessment and authentication of seafood through DNA barcoding. In: Trivedi S, Ansari AA, Ghosh SK, Rehman H (eds) DNA barcoding in marine perspectives. pp 63–70. https://doi.org/10.1007/978-3-319-41840-7_4

Sander PM (2012) Reproduction in early amniotes. Science 337(6096):806–808

Savolainen V, Cowan RS et al (2005) Towards writing the encyclopedia of life: an introduction to DNA barcoding. Philos Trans R Soc Lond B Biol Sci 360(1462):1805–1811

Smith MA, Woodley NE et al (2006) DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). Proc Natl Acad Sci U S A 103(10):3657–3662

Smith MA, Wood DM et al (2007) DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. Proc Natl Acad Sci U S A 104(12):4967–4972

Trivedi S, Affan R, Alessa A, Ansari AA, Dhar B, Mahadani P, Ghosh SK (2014) DNA barcoding of red sea fishes from Saudi Arabia – the first approach. DNA Barcodes 2:17–20

Trivedi S, Aloufi AA, Ansari AA, Ghosh SK (2015) Molecular phylogeny of oysters belonging to genus *Crassostrea* through DNA barcoding. J Entomol Zool Stud 3(1):16–21

Trivedi S, Rehman H, Saggu S, Pannerselvam C, Abbas ZK, Ahmad I, Ansari AA, Ghosh, SK (2016a) DNA barcoding in the marine habitat: an overview. In: Trivedi S, Ansari AA, Ghosh SK, Rehman H (eds) DNA barcoding in marine perspectives. pp 3–28. https://doi.org/10.1007/978-3-319-41840-7_1

Trivedi S, Ansari AA, Rehman H, Saggu S, Abbas ZK, Ghosh SK (2016b) DNA barcoding as a molecular tool for the assessment of plant biodiversity. In: Ansari AA, Gill SS, Abbas ZK, Naeem M (eds) Plant biodiversity: monitoring assessment and conservation. CABI, Wallingford, pp 419–428

Trivedi S, Aloufi AA, Ansari AA, Ghosh SK (2016c) Role of DNA barcoding in marine biodiversity assessment and conservation: an update. Saudi J Biol Sci 23(2):161–171

Trivedi S, Aloufi AA, Rehman H, Saggu S, Ghosh SK (2016d) DNA barcoding: tool for assessing species identification in Reptilia. J Entomol Zool Stud 4(1):332–337

Vargas SM, Araujo FC et al (2009) DNA barcoding of Brazilian sea turtles (Testudines). Genet Mol Biol 32(3):608–612

Wheeler QD, Raven PH et al (2004) Taxonomy: impediment or expedient? Science 303(5656):285

Wolinsky H (2012) Wildlife forensics. Genomics has become a powerful tool to inform conservation measures. EMBO Rep 13(4):308–312

Hippophae rhamnoides (Sea Buckthorn): A High-Altitude Medicinal and Adaptogenic Plant—Molecular Characterization and Bar-Coding

Shweta Saxena, Om Prakash Chaurasia, and Ratan Kumar

Abstract Medicinal plants have been used as herbal medicine by mankind since time immemorial in various traditional medicines, including Ayurveda. Sea buckthorn (*Hippophae* spp. L.) is a high-altitude-growing medicinal plant belonging to the family Elaeagnaceae. It is also known as gold mine of Himalayas. Due to genetic biodiversity of the plant, variations in phytochemicals responsible for biological activities do occur. To avoid these variations, proper identification of sea buckthorn plant is required to be done before using its product to carry out any scientific investigations or prepare any health product. Traditionally taxonomic classification of genus *Hippophae* is based on different morphological and biochemical features. These methods alone can be affected by many factors, leading to subtle and ambiguous results. *Hippophae* species are sometimes misidentified because of the similarities in vegetative morphology.

Hence, for more accurate plant identification and characterization and discrimination between medicinal and non-medicinal *Hippophae* species, powerful tools such as molecular markers and bar-coding are used. DNA bar-coding is a more recent, powerful, and accurate tool used for the identification of medicinal species of *Hippophae*. The developed markers, by using bar-coding techniques, will help to understand how genetic variability within *Hippophae* species is translated into incredibly high adaptability of its different species to abiotic stress and biotic pressure. This knowledge can be especially useful for efficient engineering of highly productive cultivars with wide range of adaptations to different environments of cultivation. The generated markers could also distinguish this taxon from possible adulterants.

Keywords Sea buckthorn · High altitude · Medicinal plant · Molecular markers · Bar-coding

S. Saxena · O. P. Chaurasia
Defence Institute of High Altitude Research, Defence Research and Development Organisation, Leh-Ladakh, Jammu & Kashmir, India

R. Kumar (✉)
Indian National Academy of Stress Sciences, Delhi, India

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_26

1 Sea Buckthorn: A High-Altitude-Growing Plant

Sea buckthorn (*Hippophae* spp. L.) is an ecologically and economically important plant that belongs to the family Elaeagnaceae. The word *Hippophae* has been derived from a Latin word *hippo* meaning horse and *phaos* which means "shine." In Greece, sea buckthorn leaves and twigs were used to feed animals which resulted in weight gain and shining coat, especially in horses. It has silvery deciduous leaves and colorful red, orange, or yellow berries that remain on the shrub throughout the winter. The plant is hardy, and it can withstand extreme temperatures from −40 °C to 40 °C. Sea buckthorn leaves are small, cuticle is thicker, and crib texture is well developed, and leaf back is densely covered with scales and star hair to cover stoma. The root nodules in sea buckthorn roots have symbiotic association with nitrogen-fixing bacterium belonging to the genus *Frankia*. The plant is found in China, Russia, India, Nepal, Bhutan, Pakistan, Afghanistan, Mongolia, Kazakhstan, Hungary, Switzerland, Romania, Germany, France, Britain, Finland, Sweden, and Norway. Several countries including India have been studying this amazing plant.

Sea buckthorn is a dwarf-to-tall (3–15 ft), branched, and thorny deciduous shrub known locally in India as "tsermang" and had been found growing wild at altitude, 7000–15,000 ft (2590–4175 m, above mean sea level). The exact number of species in the genus *Hippophae* is still unclear; however, there are considered to be seven species. The seven recognized *Hippophae* species are *H. rhamnoides* Linn., *H. tibetana* Schltdl., *H. salicifolia* D. Don, *H. goniocarpa* Y.S. Lian, *H. gyantsensis* (Rousi) Y.S. Lian, *H. neurocarpa* S.W. Liu & T.N. He, and *H. litangensis* Y.S. Lian & X.L. Chen ex Swenson & Bartish. Three species, i.e., *Hippophae rhamnoides*, *Hippophae salicifolia*, and *Hippophae tibetana*, have been described growing naturally in India in cold desert regions and other high-altitude Himalayan regions, comprising the states of Himachal Pradesh, Ladakh in Jammu and Kashmir, Uttarakhand, Sikkim, and Arunachal Pradesh in India, of which *Hippophae rhamnoides* L. ssp. *turkestanica* is the major one. In India, *H. rhamnoides* is distributed in Ladakh region of Jammu and Kashmir; Kukumseri, Lahaul, Kaza, and Tabo in Himachal Pradesh; and Nathu La in Sikkim. The *H. salicifolia* is mainly found in Uttarakhand in large areas at high-altitude regions (>2000 m). In India *H. salicifolia* is also found in Sikkim and distributed at high-altitude height from 2430 to 3660 m. *H. tibetana* is predominant in the Zanskar valley of Ladakh. *H. rhamnoides* subspecies *turkestanica* is a predominant species that grows on riversides and ravines in these valleys that are characterized by harsh weather and tough terrain (Stobdan et al. 2011). In India the average rainfall in these sea buckthorn-growing regions fluctuates from 50 to 700 mm per year. The mean temperature in Leh and Lahaul and Spiti regions shows considerable variation throughout the year (a maximum of 27 °C during summers and minimum of −30 °C in winters).

The male bud consists of 4–6 apetalous flowers, which produce wind-distributed pollen, whereas the female bud usually consists of one single apetalous flower with one ovary and one ovule. The female plants produce berrylike fruit 6–9 mm in diameter, soft, juicy, and rich in oils. The berries generally consist of pulp (68%), seed (23%), and peel (7.75%). The ripe fruit has been reported to be a source of

exceptionally high contents of vitamins (A, C, E, and K), carotenoids, flavonoids, and organic acids. Fruit juice is rich in sugar, organic acids, amino acids, essential fatty acids, phytosterol, flavonoids, vitamins, and mineral elements. Concentration of vitamins A, B, and C is much higher than other fruits and vegetables such as carrot, tomato, orange, etc. Sea buckthorn leaves are rich in flavonoids, tannins, and triterpenes (Kallio et al. 2002a; Stobdan et al. 2011). The presence of these antioxidant vitamins in high quantity indicates its strong antioxidant property.

Sea buckthorn leaf is a rich source of protein and antioxidant including tannins. Sea buckthorn leaves contain many nutrients and bioactive substances such as carotenoids, free and esterified sterols, triterpenols, and isoprenols. It contains approximately 15–20% proteins. Leaves of female and male plants are reported to contain an average of 21.1 and 20.6 g protein/100 g dried leaf, respectively. Leaf protein content peaks between the middle of July and early August and began to decrease significantly in the middle of August. Leaf tannin content peaks in the middle of July and began to decrease significantly in the middle of August. Flavonoid content in leaves ranges from 312 to 2100 mg/100 g of air-dried leaves (Stobdan et al. 2010, 2013).

Sea buckthorn has been used in traditional Tibetan and Mongolian medicine system for centuries and described in Ayurveda, one of the most ancient systems of traditional medicine, which originated in India about 5000 years ago and which heavily relies on the plant-derived therapeutics. Sea buckthorn is identified in Ayurvedic medicinal plant as *Amlich* or "Badriphal" (Sharma and Chunekar 1998) and "Amalvetas." The plant is known to Indian researchers for a long time and has been evaluated for its medicinal properties. The Indian scientists discovered and experimented on sea buckthorn as early as 1962, for its antitumor activity (Ambaye et al. 1962).

2 Health Applications

The term "medicinal plant" includes a wide range of plant species that have therapeutic properties and have been used as herbal medicine by mankind. Medicinal plants have a rich source of the specialized metabolites which are used in drug processing. Sea buckthorn (*Hippophae* spp. L.) is also a medicinal plant and has a rich history of use in treating numerous medical conditions.

China listed sea buckthorn in the pharmacopoeia in 1977. In India Ladakh region, even today, *amchies* (local traditional doctor) often prescribe preparations from sea buckthorn for treatment of common problems like indigestion, throat infection, gynecological problems, ulcer, gastritis, bronchitis, acidity, diarrhea, hypertension, blood disorder, fever, tumor, gallstone, cough, cold, food poisoning, etc. Potential health benefits of sea buckthorn may be due to its high nutrient content and antioxidant activity of various parts, viz., fruit, seed, and leaves.

Scientifically evaluated pharmacological actions of sea buckhorn are multiple such as antiplatelet aggregation, antioxidant and antibacterial, antiulcer, anti-

inflammatory, anticancer, hepato-protective, cardioprotective, cyto-protective, tissue regeneration, antitumor, anti-radiation, immunomodulatory, antihypertensive, wound healing, radioprotective, adaptogenic, anti-dermatitis, anti-atherogenic, hypoglycemic, etc. (Alam 2004; Geetha and Gupta 2011). A number of health products are being manufactured from sea buckthorn.

3 Adaptogenic Activity for High-Altitude Acclimatization

The high-altitude environment consists of various stressors, viz., hypoxia, cold, humidity, solar radiation, ultraviolet radiation, and ionizing radiation. The major stress stimuli are hypoxia and cold. Sojourners who are unable acclimatize successfully in such a stressful environment suffer from high-altitude-induced diseases such as acute mountain sickness, high-altitude pulmonary edema, high-altitude cerebral edema, retinal hemorrhages, chronic mountain sickness, systemic hypertension, frostbite, etc. (Heath and Williams 1980). But not all the individuals suffer from stress-induced maladies. Only some persons who are unable to acclimatize successfully suffer from such diseases. This indicates that genetic factors do play an important role in acclimatization and performance of the organism during exposure to adverse climatic conditions (Kumar et al. 2003, 2004; Saxena et al. 2005).

The word and concept of an "adaptogen" are a relatively new way of describing a type of remedy commonly found in traditional medicine systems. An adaptogen has "non-specific" activity and acts by increasing resistance of the organism to a broad spectrum of adverse biological, chemical, and physical factors. They help regulate or normalize organ and system function within the organism and are relatively nontoxic to the recipient. There are various herbal plant products known to have "adaptogenic" properties and used in several countries, for example, *Panax ginseng* (ginseng root) and *Eleutherococcus senticosus* (Siberian ginseng root). Some of the Rasayana, viz., *Withania somnifera* (ashwagandha root), *Ocimum sanctum* (holy basil herb), *Emblica officinalis* (amla fruit), *Asparagus racemosus* (shatavari), etc., used in Ayurveda also possess adaptogenic properties. The focal basis of Rasayana is accelerated and appropriated nutrition leading to improved biological competence of the body (Mishra 2011).

It was hypothesized that plant growing in adverse environmental condition acquires biomolecules which help them to sustain in such type of harsh climate. Such adaptogenic biomolecules, derived from plants, might be of immense importance in improving the human acclimatization in similar conditions (Divekar et al. 1996). The use of such plant products may help sojourners who are unable to acclimatize successfully in stressful environment including high altitude and suffer from high-altitude-induced diseases (Heath and Williams 1980). Oxidative stress, inflammation, altered cellular metabolism, altered immune functions, and shift in

fluid compartments are thought to play an important role in causing high-altitude-induced maladies. Sea buckthorn (*Hippophae rhamnoides* L.), growing in India at high altitude, also possesses antioxidant, anti-inflammatory, hepato-protective, cardioprotective, cyto-protective, tissue regeneration, anti-radiation, immunomodulatory, and anti-atherogenic properties. Sea buckthorn leaves were found to possess more potent activity (Geetha and Gupta 2011). Hence, adaptogenic activity of sea buckthorn leaves was examined.

3.1 Anti-stress Activity and Adaptogenic Activity

Oral administration of single-dose (100 mg/kg body weight) extract of sea buckthorn dry leaves extracted with 70% ethyl alcohol was found to possess anti-stress activity (Geetha et al. 2002). Sea buckthorn leaf aqueous extract was screened for its anti-stress and adaptogenic activity in rats using cold (5 °C)–hypoxia (428 mmHg)–restraint (C–H–R) animal model (Ramachandran et al. 1990). This animal model of multiple stressful situations was similar to conditions present at high altitude (hypoxia and cold) along with restraint stress. The maximal effective adaptogenic dose of the extract was observed to be 100 mg/kg body weight, administered orally, 30 min prior to cold (5 °C)–hypoxia (428 mmHg)–restraint (C–H–R) exposure (Saggu et al. 2007). The LD_{50} of the extract in rats was observed to be >10 g/kg body weight. Hence sea buckthorn leaf aqueous extract was an effective and safe preparation capable of enhancing performance in situations like high altitude (Saggu et al. 2007). The extract was also found to be free of any heavy metal toxicity (Saggu et al. 2006).

The mechanism of anti-stress adaptogenic activity of sea buckthorn leaf extract administered in rats at a dose of 100 mg/kg body weight during C–H–R exposures was also examined (Saggu and Kumar 2007a). Sea buckthorn leaf extract exerted its adaptogenic activity by improving aerobic metabolism. Sea buckthorn aqueous leaf extract treatment in animals maintained blood glucose levels and also restricted the rise in serum free fatty acids (FFA) and lactate dehydrogenase (LDH) levels, in comparison to controls, in attaining Trec 23 °C during C–H–R exposure. It indicated sparing of carbohydrates and more efficient utilization of FFA for energy production and better maintained cell membrane permeability (Saggu and Kumar 2008). Sea buckthorn leaf aqueous extract treatment also helped to reduce oxidative stress in the liver and muscle of rats during C–H–R exposure and post-stress recovery (Saggu and Kumar 2007b).

Though the aqueous fruit extract of sea buckthorn (*Hippophae rhamnoides* L.) also showed adaptogenic activity during C–H–R exposure, leaf extracts were more potent in comparison to fruit extract. The maximal effective adaptogenic dose of the fruit aqueous extract was observed to be 75 mg/kg body weight (Tulsawani et al. 2010). The evaluated extract was also safe when administered in either acute doses of 1 g/kg and 2 g/kg for 14 days or 75 mg/kg and 500 mg/kg for 30 days.

3.2 *Trans-vascular Leakage of Fluid*

Hypoxia is a major pathogenic factor for alterations in fluid compartments including induction of vascular leakage in the brain, an important factor for high-altitude cerebral edema. Pretreatment of animals with sea buckthorn seed oil significantly improved hypoxic tolerance and restricted the hypoxia-induced increase in fluorescein dye leakage in the brain suggesting protection against hypoxia-induced trans-vascular leakage (Purushothaman et al. 2008).

Fluid shift in pulmonary compartment is responsible for high-altitude pulmonary edema. The potential of sea buckthorn (*Hippophae rhamnoides* L.) leaf extract was examined in curtailing hypoxia-induced trans-vascular permeability in the lungs by measuring lung water content and leakage of fluorescein dye into the lungs and further confirmation by quantitation of albumin and protein in the bronchoalveolar lavage fluid. In extract-treated animals, hypoxia-induced vascular permeability decreased as evidenced by decreased water content and fluorescein leakage in the lungs and decreased albumin and protein content in the bronchoalveolar lavage fluid. The sea buckthorn leaf extract also attenuated hypoxia-induced increase in the levels of pro-inflammatory cytokines and decreased hypoxia-induced oxidative stress by stabilizing the levels of reduced glutathione and antioxidant enzymes. It provided significant protection against hypoxia-induced pulmonary vascular leakage (Purushothaman et al. 2011).

4 Genetic Biodiversity of Sea Buckthorn (*Hippophae rhamnoides* L.)

Genetic diversity enables plant species to have a greater potential for the evolution of new combinations of genes and, subsequently, a greater capacity for evolutionary adaptation to different environmental conditions. Genetic diversity is the basis for plant adaptation, evolution, and breeding.

Sea buckthorn is extremely variable in height, from a small bush less than 50 cm to a tree more than 20 m high. There is variation in growth rhythm, hardiness, and height according to the geographic distribution, i.e., the higher the latitude, the shorter the growth period and plant height. The total genetic variance in *Hippophae rhamnoides* (subspecies plus population) varied from 50% to 84% of total variance for all characters. Later growth cessation was correlated with longer growth period, taller plants, more severe frost, and winter damage. Strong cline variation showed that the higher the latitude, the earlier the growth cessation, and the shorter the growth period and plant height, the more hardy the population (Yao and Tigerstedt 1994, 1995). In Russia, Mongolia, and Germany, thornless or nearly thornless cultivars have been bred. Genetic diversity in sea buckthorn provides a good opportunity for plant breeding and selection, while variations in the growth rhythm,

height, and hardiness provide guidelines for seed and plant transfer as well as plant introduction.

Cold hardiness is a very complex trait that is controlled by many genes. Cold acclimation increases the concentration of sugars and levels of dehydrin which, in turn, result in elevated freezing tolerance. Sea buckthorn has cold hardiness to temperatures ranging from −30 to −35 °C and −40 to −45 °C (Tang and Tigerstedt 2001). With the increase of cold stress intensity, the photosynthesis rate, transpiration rate, stomatal conductance in leaves, and contents of abscisic acid and indole acetic acid in roots decrease, but water-use efficiency, abscisic acid, and zeatin riboside in leaves increase. 32 out of 39 differentially expressed proteins under low-temperature stress have been identified, and their functions were mainly involved in metabolism, photosynthesis, signal transduction, anti-oxidative systems, and posttranslational modification (He et al. 2016). The changed protein abundance and corresponding physiological-biochemical response shed light on the molecular mechanisms related to cold tolerance in cold-tolerant plants and provide key candidate proteins for genetic improvement of plants.

Considerable genetic variations are observed among berry quality-related traits such as berry size, sugar components, vitamin C, and titrable acidity among subspecies and hybrid groups. Different origins also varied for rhythm of berry maturation. During berry maturation, concentrations of sugar components, especially glucose and ratio of sugar to acidity, increase, while concentration of vitamin C, titrable acidity, and berry hardiness decreases. In crosses between subspecies, berry size, vitamin C, and titrable acidity have shown intermediate inheritance (Tang 2002).

The oil content and fatty acid composition of berries from two subspecies of sea buckthorn (*Hippophae rhamnoides* L.) investigated (Yang and Kallio 2001) were found to be varying. The berries of ssp. *rhamnoides* contained a higher proportion of oil in seeds (11.3% vs 7.3%), berries (3.5% vs 2.1%), and seedless part (2.8% vs 1.7%) than the berries of ssp. *sinensis*. More linoleic acid (40.9% vs 39.1%) and less α-linolenic acid (26.6% vs 30.6%) were found in the seed oil of ssp. *sinensis* than in the seed oil of ssp. *rhamnoides*. The proportion of palmitoleic acid was higher in the oil of berries of ssp. *rhamnoides* than the berries of ssp. *sinensis* (26.0 vs 21.5%).

Kallio et al. (2002b) also observed that seeds of ssp. mongolica were a better source of tocopherols and tocotrienols than those of ssp. sinensis (287 vs 122 mg/kg). The proportions of linoleic acid and palmitic acid in berries and seeds in glycerophospholipids were higher in ssp. mongolica than in ssp. sinensis. It showed that compositional differences between the two subspecies should be considered when the berries are bred and exploited for nutritional purposes.

In another study, vitamin C, tocopherols, and tocotrienols in berries of wild and cultivated sea buckthorn (*Hippophae rhamnoides* L.) of different origins and harvesting dates were determined with HPLC (Kallio et al. 2002a). Wild berries of ssp. sinensis, native to China, contained 5–10 times more vitamin C in the juice fraction than the berries of ssp. rhamnoides from Europe and ssp. mongolica from Russia (4–13 vs 0.02–2 g/L juice). Genetic background and berry harvesting date were two primary factors determining the vitamin C content in the berries. The seeds

of ssp. sinensis contained less tocopherols and tocotrienols (average 130 mg/kg) compared with seeds of ssp. rhamnoides (average 290 mg/kg) and *mongolica* (average 250 mg/kg). The fruit flesh of sinensis berries had contents of tocopherols and tocotrienols 2–3 times higher than those found in the other two subspecies (120 mg/kg vs 40 mg/kg in rhamnoides and 50 mg/kg in mongolica).

17 Indian natural population of sea buckthorn (*Hippophae rhamnoides* L.), which comprised of 187 plants from Trans-Himalaya, were studied to find out the variability and genotypic and morphometric effect on fruit oil content (Korekar et al. 2013). The soft part of the fruit was found to be rich in oil content ranging from 3.16 to 17.3%. A variation of 1–5.5-folds in oil content among the examined fruit underlines the important role played by genetic background of the plant for determining oil content in sea buckthorn fruit. The overall mean oil content was 9.74%. Oil content in berry was significantly correlated with fruit weight, fruit width, tree habit, seed weight, seed length, leaf area, number of leaf per 10 cm^2, secondary branch apex, and mature stem color.

Seventeen natural populations of Indian sea buckthorn also showed variability (Korekar et al. 2014) in total phenolic content (964–10,704 mg gallic acid equivalent/100 g), free radical-scavenging activity in terms of inhibitory concentration (IC_{50}) (ranged from 0.7 to 9.1 mg/mL), ferric reducing antioxidant potential (180–1355 $FeSO_4.7H_2O$ μg/mL), and ascorbic acid and carotenoid content (56–3909 mg/100 g and 0.1–14.4 mg/100 g, respectively).

So, the presence of various biomolecules in herbal products of *Hippophae rhamnoides* L. has been found to vary depending upon species, the origin of place of the plant, as well as harvesting time. It means that proper identification of subspecies of the plant is important because they differ in biomolecules and so in their medicinal effects also.

Hence, proper required identification of sea buckthorn plant has to be done before using its products to carry out any scientific investigation or prepare any health product. Otherwise a large variation in experimental results of scientific studies will be observed, and even prepared health products will also prove ineffective due to batch-to-batch variation in active biomolecules of herbal product used, on account of unidentified plant biodiversity.

5 Different Methods Used for Identification and Characterization of Sea Buckthorn (*Hippophae rhamnoides* L.)

Many related herbs have different common names and vice versa. For the protection of consumers, authentication of medicinal plants is a critical issue. Ideally, authentication should occur from the harvesting of the plant material to the final product. Unfortunately there is no single or superior method to assure 100% authentication during the entire process, but the goal can be achieved through the application of a

variety of different methodologies. Various methods have been used to identify medicinal plants starting from old days up to recent times. Some of them are parataxonomy and inventory interview (Frank-Thorsten 2004), inventory interview (Silva et al. 2014), image processing and matching with existing database, morphological characterization, chemical and biochemical characterization, molecular markers (Agarwal et al. 2008; Chen et al. 2014), and DNA bar-coding. The latter are more suitable for identification as deoxyribonucleic acid (DNA), in contrast to ribonucleic acid, is a stable macromolecule that is found in all tissues, including the dead ones.

The whole process starts with good voucher specimens that act as reference material and to prove chain of custody. Macroscopic and microscopic examinations can be used as rapid and inexpensive identification techniques. Chemical analysis is by far the best method for the detection of contaminants and can be an excellent method for plant identification. Each of these methodologies has limitations, and more analytical methods are needed to assist in the authentication process. Molecular biology offers an assortment of techniques that can be very useful for authentication of medicinal plants. Molecular marker techniques provide valuable data on diversity through their ability to detect variation at the DNA level and have been routinely employed as they are easy, rapid, and reliable to perform. Many DNA-based methods have been used for detecting and differentiating plant materials from their contaminants, viz., random amplified polymorphic DNA (RAPD), polymerase chain reaction (PCR)-restriction fragment length polymorphism (RFLP), high-resolution melting analysis (Jaakola et al. 2010), DNA bar-coding (Techen et al. 2014), etc. Among these, DNA bar-coding is a comparatively recent technique, which uses DNA sequences from a small fragment of the genome to identify organisms (Hebert et al. 2003). Furthermore, bioinformatic approaches have been of immense help for the in-depth understanding of the plant systems in various ways.

5.1 Morphological Characterization

Morphological traits of seeds and fruits are considered as diagnostic characters of *H. rhamnoides* to reconstruct the relationship among the collections made from different regions and to test whether there is a significant association between the morphological characters measured and its environment. Traditionally taxonomic classification of genus *Hippophae* is based on different morphological and biochemical features. However, these methods alone can be affected by many factors, leading to subtle and ambiguous results.

For identification, the simplest, most direct, and least expensive is macroscopic identification. Morphological characteristics can be observed by macroscopic and microscopic methods (identification by shape, color, and texture) with a hand lens or dissecting microscope, as well as organoleptic characteristics (e.g., color, taste, mouth feel, and odor) that are useful for confirming plant identity. The medicinal plants are identified using structural features, namely, height, shape, size of leafy

part, flowers, fruits, and branching patterns. A careful study of the morphological traits of sea buckthorn (*Hippophae* spp.), coupled with unweighted hierarchical clustering, has the potential to provide results on par with RAPD analysis (Mathew et al. 2007), making it possible to avoid expensive and tedious procedures.

Plant recognition system also uses number of petals, bloom color, inflorescences, shape of blooms, and shape of leaves to identify a plant. It also proposes a medicinal plant database for indexing and retrieving the medicinal plant characteristics.

5.2 Chemical and Biochemical Characterization

Various available scientific tools, e.g., TLC, HPTLC, HPLC-UV, and HPLC-MS, are used for chemical profiling of the plant with respect to its chemicals present in it (Soni et al. 2013). N-glycoproteome analysis, including protein disulfide isomerase, by con A affinity chromatography from seedling of *Hippophae rhamnoides* was recently attempted. A total of 48 spots were detected, and 10 nonredundant proteins were identified by MALDI-TOF/TOF (Sougrakpam and Deswal 2016). Biochemical analysis of sea buckthorn berries has revealed a wide range of variations in vitamin C, carotene, flavonoid, and vitamin E concentrations among individual genotypes and populations. Biochemical markers like isozymes revealed polymorphisms in 25 populations of *Hippophae* at the protein level and have been used for studying genetic variation within a large number of species (Yao and Tigerstedt 1993).

5.3 Molecular Markers

Morphological characterization is a conventional technique used for evaluating the plant genetic diversity, whereas molecular techniques are relatively new and more effective in this context. Molecular markers are far more efficient than morphological and biochemical markers in unequivocally establishing the phylogeny and taxonomy of sea buckthorn in addition to their role in breeding programs and germplasm characterization.

With recent advances in molecular biology, DNA molecular markers are being increasingly employed in plant genetic and breeding research. Many PCR-based DNA markers have been developed, including random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR), expressed sequence tag (EST), and inter-simple sequence repeats (ISSRs). Among these DNA markers, RAPD and ISSR are relatively simple to use and are highly effective in plant fingerprinting and phylogenetic studies, which require no prior knowledge of sequence information. Sequence-characterized amplified region (SCAR) markers with designed primers significantly improve the reproducibility and reliability of PCR-based assays (Hernández et al. 1999). They are therefore considered to be more suitable molecular markers for crop breeding and genetic

analysis. Furthermore, SCAR markers have relatively long sequences and can produce more specific and prominent band signals (Chowdhury et al. 2001).

Different types of genetic markers have been used to assess phylogenetic relationships and diversity in sea buckthorn: initially, isozymes (Yao and Tigerstedt 1993) and later fatty acid methyl ester (FAME) analysis and nucleic acid-based techniques, such as RAPD (random amplified polymorphic DNA) (Ercili et al. 2008; Ruan et al. 2004), inter-simple sequence repeats (ISSRs), amplified fragment length polymorphism (AFLP), sequence-related amplified polymorphism, and simple sequence repeat (SSR).

5.3.1 Random Amplified Polymorphic DNA

Of the various molecular techniques available, random amplified polymorphic DNA (RAPD) is widely used for estimating genetic diversity and relationships in plant populations. RAPD is a type of PCR reaction, but the segments of DNA that are amplified are random. No knowledge of the DNA sequence for the targeted gene is required, as the primers will bind somewhere in the sequence, but it is not certain exactly where. This makes the method popular for comparing the DNA of biological systems that have not had the attention of the scientific community or in a system in which relatively few DNA sequences are compared (it is not suitable for forming a DNA data bank). RAPD markers are decamer (ten nucleotide length). DNA fragments from PCR amplification of random segments of genomic DNA with single primer (8–12 nucleotides) are able to differentiate between genetically distinct individuals. RAPD has been used to analyze the genetic diversity and characterize and trace the phylogeny of diverse plant and animal species. The genetic identification of sea buckthorn using RAPD data has been performed in many countries including China (Sun et al. 2006) and India (Singh et al. 2006).

RAPD analysis was used to investigate the genetic diversity and population genetic structure of 13 natural populations of the sea buckthorn subspecies *sinensis* (*Hippophae rhamnoides* ssp. *sinensis*). 15 primers amplified 107 reproducible bands, with 95 (88.79%) being polymorphic. The gene diversity within population was 0.168, considerably lower than that of tree species and most perennial, outcrossing species but higher than that of annual or short-lived species. The genetic diversity analysis Gst value showed that 18.3% of the total genetic variation resided among populations, a little lower than that of outcrossing species. The low genetic differentiation among populations in ssp. sinensis may be attributed to the long-distance dispersal of seeds facilitated by birds, in addition to its characteristics of outcrossing, wind pollination, and widespread distribution. The studies have shown a low level of genetic distance suggesting a high level of gene flow or genetic variability which helps the species to evolve. No association between genetic distance and geographical distribution was found (Sun et al. 2006). The ssp. from Northern Europe was studied using RAPD, and more genetic variation was observed within the populations than among the populations (Bartish et al. 2000).

The differences in the fatty acid profile illustrated the chemotaxonomic relationships among the genotypes and demonstrated the importance and potential of fatty acids in delimitation of sea buckthorn genotypes. Fatty acid methyl ester (FAME) and nucleic acid RAPD profiles were used to examine biochemical and genetic relationships between ten selected sea buckthorn genotypes found in the same area of Eastern Anatolia (Ercili et al. 2008). Fatty acid composition of sea buckthorn berries was determined using gas chromatography. Fatty acid results showed that there were differences between genotypes in both the percent and presence of fatty acids in the berries. RAPD analysis showed that there were distinct genetic differences between the genotypes. All ten genotypes of *H. rhamnoides* were clearly differentiated based on a dendrogram constructed. RAPD data also showed that the ESB4 genotype was in a group distinct from the others (Ercili et al. 2008).

RAPD markers have also been used to determine the genetic variation between and within native populations and to measure relatedness between populations of the genus *Hippophae*. RAPD (random amplified polymorphic DNA) analyses were conducted with nine decamer primers, which together yielded 219 polymorphic markers. Sixteen fixed RAPD markers, i.e., markers that either occurred in all plants of a population or were absent from all plants, were found. Several of these markers were useful for analysis of interspecific relationships, whereas others were considered as taxon-specific markers. Amount and distribution of genetic variability varied considerably between species. Partitioning of molecular variance within *H. rhamnoides* supported findings that a considerable part of the total variance resides among subspecies (59.6%). Within-population variability also differed considerably (Bartish et al. 2000).

5.3.2 Analysis of Indian Sea Buckthorn spp.

Indian sea buckthorn (*Hippophae rhamnoides* L.) species growing in Ladakh (altitude range of 8000–13,500), a cold desert in the Trans-Himalayan range of 32°N to 76°E, is locally called as "tsermang," and the fruits are called as "tsestalulu." The region is characterized by extreme climatic conditions like high wind velocity, high rate of soil erosion, and extreme temperatures, i.e., −30 °C in winter to +35 °C in summer. Genetic variability of this Indian sea buckthorn (*Hippophae rhamnoides* L.) species was examined by RAPD analysis in 17 morphotypes found in Nubra valley (Singh et al. 2006). Young leaves were collected during August and genomic DNA was isolated. DNA was further purified and utilized in PCR reactions. A set of 20 random decamer primers selected were tested across all 17 morphotypes. There were 18.86 amplicon levels per primer, of which 16.80 were polymorphic indicating higher variability among *H. rhamnoides* population. Population similarity coefficient matrix based on shared banding pattern showed wide differences. Polymorphism revealed by RAPD markers serves as a dominant Mendelian marker.

RAPD polymorphisms were analyzed with a phenatic distance measure (Jaccard coefficient) from which a dendrogram was constructed using median method, providing an indication of diversity present within the *H. rhamnoides* population. Clustering analysis clearly showed five major groups in the natural population of

H. rhamnoides. One of the important observations of the study was that none of the morphotypes showed a similarity more than 25%, indicating high level of genetic difference between morphotypes. The RAPD analysis of all the 17 morphotypes revealed low level of genetic distance suggesting high level of gene flow (Singh et al. 2006). Genetic variability in the natural sea buckthorn population indicated the presence of subspecies in the region, which needs to be confirmed by taxonomists.

Another study (Srihari et al. 2013) examined genetic diversity of 48 sea buckthorn genotypes collected from 16 locations in 4 diverse geographical regions of Himachal Pradesh and Jammu and Kashmir states of northwestern Himalayan region of India, using random amplified polymorphic DNA (RAPD) along with other molecular markers. Sufficient genetic differentiation within populations and a low level of variation among populations were observed.

5.3.3 Amplified Fragment Length Polymorphism Markers

Various methods have been developed for the identification and typing of prokaryotic and eukaryotic organisms at the DNA level. AFLP analysis belongs to the category of selective restriction fragment amplification techniques, which are based on the ligation of adapters (i.e., linkers and indexers), to genomic restriction fragments followed by a PCR-based amplification with adapter-specific primers (Savelkoul et al. 1999). AFLP analysis is an established and broadly applicable genotyping method with high degrees of reproducibility and discriminatory power. Both RAPD and AFLP markers are useful for studying relationships among DNA sequences, independently of their location in the chromosomes or of particularities in their nucleotide sequence, but the higher multiplex ratio of AFLP markers makes them more suitable for distinguishing between closely related genotypes, such as different clones within a given cultivar.

AFLP markers have been used to study genetic relationships in many plant species. The AFLP markers were also found useful for fingerprinting and detecting genetic relationships among sea buckthorn cultivated varieties. In a study, AFLP markers were used for analyzing DNA fingerprint patterns and genetic relationships among 15 sea buckthorn cultivated varieties from China, Russia, and Mongolia, which were mainly and widely planted and extended varieties in China (Ruan and Li 2005). In a Pakistani study, 25 sea buckthorn ecotypes were analyzed. In this study, the leaf samples were taken from wild sea buckthorn populations from Skardu, Hunza Gilgit, Hussainabad, Murtazaabad, Nagar, Shigar, Sadpara, Khaplu, Sust, and Jaglot areas of Northern Pakistan (Shah et al. 2009). Phylogenetic distance estimated revealed that the ecotypes expressed common heritage for their phylogenetic relationship with a considerable genetic diversity among them as well.

Results demonstrated that AFLP markers are useful for fingerprinting and detecting genetic relationships among sea buckthorn cultivated varieties. The results can be used as guidelines for improving germplasm collection and breeding.

5.3.4 Sex Determination

Since sea buckthorn is dioecious, in the commercial production of sea buckthorn berries, the quality of the female plants is critical, whereas only about 10% male plants are needed in the field to produce large amounts of fertile pollen. Gender is most often genetically determined in dioecious plants, either by distinguishable sex chromosomes or by alleles at one or several loci on non-distinguishable chromosomes (Durand and Durand 1990). Unfortunately, gender of sea buckthorn seedlings cannot be determined until flowering, which usually takes place after 3–4 years in the field. However, the mechanism governing gender determination or markers linked to sex have not been established yet in sea buckthorn. Since sea buckthorn is basically dioecious, identification of superior paternal parents is a prerequisite for achieving the breeding goals.

The random amplified polymorphic DNA (RAPD) method has been used to identify markers linked to sex determination in the dioecious species *Hippophae rhamnoides*. RAPD markers seem to be useful for genome mapping and marker-assisted selection in sea buckthorn (Persson and Nybom 1998). Korekar et al. (2012) developed robust PCR-based marker(s) to distinguish male genotypes from female genotypes early in the vegetative growth phase of *Hippophae rhamnoides* ssp. *turkestanica* from the Trans-Himalayan Ladakh region of India. DNA bulk samples from 20 male and 20 female plants each were screened with 60 RAPD primers. Two primers, OPA-04 and OPT-06, consistently amplified female-specific polymorphic fragments of 1164 and 868 bp, respectively, which were absent in the male samples. A sequence-characterized amplified region (SCAR) marker *HrX1* (JQ284019) and *HrX2* (JQ284020), designed for the two fragments, continued to amplify the female-specific allele in 120 female plants but not in 100 male plants tested. Hence, HrX1 and HrX2 are female-specific markers that can determine the sex of sea buckthorn plants in an early stage and expedite cultivations for industrial applications. This study first time reported two robust sex-linked SCAR markers in sea buckthorn.

Sea buckthorn homologous genes involved in floral development which may have role in sex determination were studied. 44 putative genes involved in sex determination (GISD) and reported in model plants were short-listed from literature survey, and 29 sea buckthorn homologous sequences were identified from available sea buckthorn genomic resources. Of these, 21 genes were found to differentially express in either male or female flower bud stages. *HrCRY2* was significantly expressed in female flower buds only, while *HrCO* had significant expression in male flowers only (Chawla et al. 2015). Information on these sex-specific expressed genes may help in elucidating sex determination mechanism in sea buckthorn.

A recent study examined the genetic diversity of *H. rhamnoides* ssp. *turkestanica* populations in Ladakh with the aim to identify a gender-specific marker using ISSR markers (Das et al. 2016). A total of 92 collections were made and the distance between any two samples was at least 10 m. Samples from the Indus Valley consisted of 12 male and female plants each. 25 males and 21 females were collected from Nubra Valley. From the Suru valley, 10 males and 12 females were harvested.

58 ISSR primers were used to characterize the genome of *H. rhamnoides* ssp. *turkestanica*, of which eight primers generated 12 sex-specific fragments specific to one or more populations. The ISSR primer (P-45) produced a fragment which faithfully segregates all the males from the female plants across all the three valleys surveyed. This male-specific locus was converted into a sequence-characterized amplified region (SCAR). Forward and reverse primers designed from this fragment amplified a 750 bp sequence in males only, thus specifying it as an informative male-specific sex-linked marker. This SCAR marker was further validated for its capability to differentiate gender on an additional collection of plants, representing three geographically isolated valleys (Nubra, Suru, and Indus) from Ladakh region of India. The results confirmed sex-linked specificity of the marker suggesting that this conserved sequence at the Y chromosome is well preserved through the populations in Ladakh region.

The genetic diversity of populations as surveyed by ISSR primers revealed 85.71% polymorphism at the population level. The dendrogram obtained divided the genotypes into three different clusters, and the distribution of male and female genotypes in all the clusters was random. Nei's genetic similarity index was in the range of 0.63–0.96. It is envisaged that the developed SCAR marker shall provide a reliable molecular tool for early identification of the sex in the commercial crop of *H. rhamnoides* ssp. *turkestanica* (Das et al. 2016). The study also signifies the applicability of ISSR markers for studying the genetic diversity and genetic relationships in different genotypes of *H. rhamnoides* ssp. *turkestanica*.

5.4 *Inter-simple Sequence Repeats and Simple Sequence Repeats*

Inter-simple sequence repeat (ISSR) markers are abundant throughout the genome and show a higher level of polymorphism than any other genetic markers. ISSRs have a number of advantages, for example, low quantities of template DNA are required, no sequence data are needed for primer construction, many informative bands are generated per reaction, and it is reliable and reproducible. Consequently, ISSRs have been widely used in marker-assisted selection, genetic diversity analysis, DNA fingerprinting, and evolution and molecular ecology. Microsatellites or simple sequence repeats (SSRs) are tandemly repeated DNA regions with motif lengths of 1–6 base pairs. By virtue of their hypervariability, abundance, reproducibility, codominant nature, and easy detection technique, SSRs are considered as one of the most efficient molecular markers which have a wide range of applications such as genetic mapping, gene tagging, and studies of genetic diversity and evolution (Varshney et al. 2005).

ISSR markers have been used widely in sea buckthorn diverse species (Ruan et al. 2004; Ruan and Li 2005) due to an ability to identify high levels of polymorphism, to

reliably differentiate between genotypes and identify markers associated with desirable traits (Ruan et al. 2009).

The genetic relationships of 51 cultivars and lines of *H. rhamnoides* ssp. *mongolica* and its hybrids with *H. rhamnoides* ssp. *sinensis* were analyzed using 13 ISSR primers, which clustered into three major groups based on 137 polymorphic markers. Associations of ISSR markers with dry pulp oil contents were also evaluated. Four ISSR markers (881_{340}, 825_{1000}, 817_{380}, and 807_{1100}) had positive association with high dry pulp oil content ($P < 0.01$) using stepwise multiple regression analysis (Ding et al. 2016).

Genetic diversity of 48 sea buckthorn genotypes collected from 16 locations in 4 diverse geographical regions of Himachal Pradesh and Jammu and Kashmir states of northwestern Himalayan region of India was analyzed using inter-simple sequence repeat (ISSR), simple sequence repeat (SSR), and MADS-box gene markers. While 21 decamer primers generated 155 loci, of which 148 (95.48%) were polymorphic with average of 7.3 loci per primer, 10 ISSR primers generated 86 loci (8.6 loci per primer), out of which 83 (96.51%) were polymorphic. Six SSRs generated 18 loci (3 loci per primer), of which 17 (94.45%) were polymorphic, whereas three SSR primers resulted into monomorphic bands. Two MADS-box primers on the other hand generated 13 loci (6.5 loci per primers) with 100% polymorphism. RAPD, ISSR, SSR, and MADS-box gene-specific primers generated amplicons with polymorphism information content (PIC) values of 0.60–0.88, 0.60–0.91, 0.23–0.73, and 0.56–0.80, respectively. Cluster analysis of genotypes was done which revealed 11 groups that followed clustering pattern more or less according their geographical origin. Sufficient genetic differentiation within populations and a low level of variation among populations were observed. Some species-specific diagnostic markers were also identified (Srihari et al. 2013).

In another study, genetic diversity among 36 genotypes of *Hippophae salicifolia* D. Don of Uttarakhand region and *Hippophae rhamnoides* ssp. *turkestanica* of Ladakh region was studied, using simple sequence repeat (SSR) DNA-PCR analysis (Sharma et al. 2015). SSR markers were used to assess the genetic diversity and inter- and intraspecific relationships of the genus *Hippophae* and to study the correlation between genetic distances and geographic distances among different genotypes of sea buckthorn.

A total of seven simple sequences repeat primers were used for detecting genetic variability in sea buckthorn seeds. SSR profile obtained by primer UTR-15 was containing a total of eight bands (0.15–0.6 kb), and among these eight bands, one was found to be polymorphic (12.5%). A dendrogram constructed based on the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) clustering method revealed two major clusters of Uttarakhand (with subclusters) and Ladakh region (no subclusters). The phenol contents of the *H. salicifolia* (particularly genotype ST 10) were found significantly higher than *H. rhamnoides* (genotype DIHAR-6). Used molecular and protein-based markers helped in the selection of superior genotypes having higher nutraceutical in Uttarakhand and Ladakh regions (Sharma et al. 2015).

In the study, although SSR-PCR and SDS-PAGE profiles were reproducible and generated several bands, the banding patterns observed with protein profiling were

not similar and discriminatory as observed with DNA and protein fingerprinting. It suggested that SSR-PCR and protein profiles which are a rapid and simple tool could be used in typing and differentiating a large number of sea buckthorn genotypes collected from different states and geographical region (Sharma et al. 2015).

Variability of ISSR molecular markers has also been used to determine association with dried-shrink disease resistance and help identify potential breeding varieties and lines of sea buckthorn. Dried-shrink disease is a dangerous pathogen that destroys sea buckthorn and halts commercial production. It decreases fruit yield by 20–40% and usually appears on plants that are at least 3 years old. 15 ISSR primers generated 346 bands, with an average of 23 bands per primer. Genetic similarities (Jaccard coefficient) between pairs of accessions ranged from 0.216 to 0.781. At a Jaccard coefficient of 0.39, the dendrogram generated with 342 polymorphic bands clustered 51 accessions of *H. rhamnoides* into five groups and one *H. salicifolia* isolate. Four ISSR markers (887_{190}, 835_{700}, 809_{290}, and 811_{280}) were found to be significantly correlated with resistance to dried-shrink disease (Ruan et al. 2009). These markers provide a potential method for breeding programs that select lineages resistant to dried-shrink disease, especially when no other genetic information such as linkage maps and quantitative trait loci are available.

The use of these ISSR markers is a potential strategy to select genotypes with high dry pulp oil content and suitable parental combinations for improvement of sea buckthorn berries.

6 DNA Bar-Coding

DNA bar-coding is a term that was first proposed by Hebert et al. (2003). DNA bar-coding (a tool for rapid species identification based on DNA sequences) and genomics (which compares entire genome structure and expression) share an emphasis on large-scale genetic data acquisition that offers new answers to questions previously beyond the reach of traditional disciplines. The DNA composition and information are unique for each species. DNA barcodes and genetic markers have a potential role to develop fingerprints. DNA barcodes consist of a standardized short sequence of DNA (400–800 bp) that in principle should be easily isolated and characterized (Savolainen et al. 2005). For DNA bar-coding, the nuclear, mitochondrial, and chloroplast DNA are used, out of which mitochondrial and chloroplast genomes are more important if the species shared intricate phylogenetic relationship. Having conserved genes, high copy number and reproducibility, and general primers makes them suitable for bar-coding (Chen et al. 2010).

Three important principles of DNA bar-coding are standardization, minimalism, and scalability. Translating this into the selection of bar-coding regions involves choosing one or a few standard loci that can be sequenced routinely and reliably in very large and diverse sample sets, resulting in easily comparable data which enable species to be distinguished from one another. Various barcode markers have been suggested which can be used or in combination for bar-coding of plants.

In 2009, the Plant Working Group of the Consortium for the Barcode of Life (CBOL) recommended a core barcode for identification across land plants, consisting of portions of two plastid coding regions, rbcL+matK, to be supplemented with additional markers as required (CBOL 2009). Some of them are:

M*atK*—formerly known as orfK is one of the most rapidly evolving coding sections of the plastid genome and plays an important role in systematic studies on the evolution of plants (Hilu and Liang 1997).

Ribulose-1, 5-biphosphate carboxylase (rbcL)—chloroplast gene rbcL is an ideal barcode in phylogenetic studies because of high degree of generality, convenience for use in amplification, its ability to distinguish species of medicinal plants at genus levels, and identification of both monocots and dicots (Reddy 2009).

trnH-psbA—beyond the core rbcL+matK barcode, the most widely used plastid bar-coding marker is the intergenic spacer trnH-psbA. Due to its high discrimination capability, the genomic region with nrITS was introduced as barcodes in plants (Kress et al. 2005).

nrITS—the internal transcribed spacers from nuclear ribosomal DNA (nrITS) are an obvious choice of a supplementary barcode in groups in which direct sequencing is possible (CBOL 2009). The generally shorter length of the target region can make routine sequencing easier than the entire nrITS, and in general, the nrITS2 region is more length-conserved than nrITS1.

DNA bar-coding will allow users to efficiently recognize known species and speed up the discovery of species yet to be found in nature. Lahaye et al. (2008) pointed out that plant DNA barcodes can be used to assess species identification in conservation biodiversity hotspots as well as hypothetically applied to monitoring the international trade in endangered species of orchids. By harnessing advances in molecular genetics, sequencing technology, and bioinformatics, DNA bar-coding allows users to quickly and accurately recognize known species and retrieve information about them.

Hollingsworth et al. (2011) gave a list and characteristics of different markers included in plant bar-coding studies. These barcode markers are given below in the list.

Characteristics of different markers that have routinely been included in plant bar-coding studies (Hollingsworth et al. 2011)

Marker	Genomic source	Type	Approx. number of GenBank accessions	Median amplicon length (bases) in completely sequenced plastid genomes	IQR amplicon length (bases)	Amplicon length range (bases)
nrITS	Nuclear	Transcribed spacers	102,684	705	683–724	407–1630

(continued)

Marker	Genomic source	Type	Approx. number of GenBank accessions	Median amplicon length (bases) in completely sequenced plastid genomes	IQR amplicon length (bases)	Amplicon length range (bases)
nrITS2	Nuclear	Transcribed spacer	111,370	494	492–506	157–670
atpF-H	Plastid	Intergenic spacer	1180	669	578–707	390–918
matK	Plastid	Protein coding	34,647	889	880–889	862–910
psbK-I	Plastid	Intergenic spacer	1241	468	444–492	112–1253
rbcL	Plastid	Protein coding	27,725	654	654–654	654–654
rpoB	Plastid	Protein coding	3341	548	548–548	536–590
rpoC1	Plastid	Protein coding	5314	616	616–616	610–622
trnH-psbA	Plastid	Intergenic spacer	23,526	509	401–617	226–934
trnL-F	Plastid	Intron and intergenic spacer	59,197	994	907–1037	201–2114
trnL (P6)	Plastid	Intron	70,811	87	83–91	51–135

7 Need for Bar-Coding of Medicinal Plants

Medicinal plants are used throughout the world, and the regulations defining their proper use, such as identification of the correct species and verification of the presence, purity, and concentration of the required chemical compounds, are widely recognized. Herbal medicines are made from vegetal drugs, the processed products of medicinal species. These processed materials present a number of challenges in terms of botanical identification. According to the World Health Organization (WHO), the use of incorrect species is a threat to consumer safety. Studies have revealed that the level of substitutions may be very high, at times up to 70%, caused intentionally (by mixing undesired substances causing adulteration) or unintentionally (improper identification of the herbal product used). So in the sold or used plant species, correct chemical compounds or phytochemicals may be absent or not present in required concentration. Such herbal product will be lacking the desired medicinal effects. The lack of authenticated medicinal plant raw drugs is a growing concern. The safety for human consumption of such

products remains unknown. Misidentification and substitutions are a reality, with confirmed reports from several countries; these issues are of high concern because they may cause fatalities among users. Cases of adulteration and contamination have led to severe illness and even death in some cases. To guarantee the quality of herbal medicines, certain steps established in the pharmacopoeias must be followed, including correct identification of the plant species, analysis of the purity, and confirmation of the presence and minimum concentration of the active ingredients [chemical marker(s)].

Accurate and fast scientific authentication/identification of the plant(s) is the key to success for the herbal drug industry. The conventional approach is to engage an expert taxonomist, who uses a mix of traditional and modern techniques for precise plant identification. However, for bulk identification at industrial scale, the process is protracted and time-consuming. It is vital to provide a single database containing information about authentic plant materials and their potential adulterants (Techen et al. 2014). DNA bar-coding, a powerful tool, offers an alternative and feasible taxonomic toolbox for rapid and robust species identification and is now being used globally to identify a plant species including endangered species. The DNA barcode consists of one or more short, standardized DNA region(s) that can be used to identify a species. For the success of DNA barcode, the barcode loci must have sufficient information to differentiate unambiguously between closely related plant species and discover new cryptic species. For herbal plant identification, matK, rbcL, trnH-psbA, ITS, trnL-F, 5S-rRNA, and 18S-rRNA have been used as successful DNA barcodes (Mishra et al. 2016).

Ayurveda, one of the most ancient systems of traditional medicine, which originated in India about 5000 years ago, heavily relies on the plant-derived therapeutics. It is estimated that about 1587 plants are used for the preparation of various kinds of Ayurvedic medicines. The lack of authenticated medicinal plant raw drugs is a growing concern in Ayurveda too, and utility of DNA bar-coding in authenticating medicinal plant raw drugs has been recognized. Ayurvedic Pharmacopoeia of India (API)—Reference DNA Barcode Library (API-RDBL) was created with high-quality and authentic *rbcL* barcodes for 374 out of the 395 medicinal plants that are included in the API. DNA bar-coding revealed that only 79% of raw drugs were authentic and the remaining 21% of samples were adulterated (Vassou et al. 2016).

8 Bar-Coding of Sea Buckthorn (*Hippophae* spp.)

In *Hippophae* (fam. Elaeagnaceae) (Shaji), 7 species and 11 subspecies have been identified worldwide. *Hippophae* species have been widely used in food, pharmaceutical, and healthcare products. *Hippophae* species are sometimes misidentified because of the similarities in vegetative morphology. Furthermore, the fruits of different species are labeled with the same name and mainly sold or used in the dried form or as powders. Therefore, different species cannot be identified by only morphological characteristics, and accurate identification methods are needed. With the advantages of high PCR amplification efficiencies, DNA sequencing success

rates, and discrimination power, DNA bar-coding has become popular with taxonomists and has gained wide acceptance as a standard and effective method in biodiversity research and conservation genetics.

Nuclear DNA data provide valuable information in phylogenetic study of plants, and the internal transcribed spacer (ITS) regions of the nuclear ribosomal DNA (nrDNA) have been shown to be a valuable source of evidence to resolve phylogenetic relationships at different taxonomic levels, in particular at intraspecific level because of the relatively rapid evolutionary rates of the ITS fragment.

The phylogenetic relationships among 15 taxa of *Hippophae* were analyzed by comparing the sequences of the internal transcribed spacer (ITS) sequences of nuclear ribosomal DNA (nrDNA). Complete sequences of the ITS regions were generated from all materials, including 15 taxa of *Hippophae* and 3 out-groups (Sun et al. 2002). The boundaries of ITS1, ITS2, and adjacent coding regions were determined. ITS sequences in *Hippophae* varied in length from 651 bp to 666 bp. The ITS sequences of three out-groups were between 646 bp and 651 bp in length. The length of ITS1 of *Hippophae* ranged from 270 bp to 282 bp, while ITS2 ranged from 218 bp to 220 bp. The amount of polymorphism observed within a taxon was found extremely low in most taxa except the two putative hybrids. The ITS sequences of two putative hybrids showed high rates of polymorphism. Comparative analysis indicated that the polymorphism site in the putative hybrids resulted from the nucleotide additivity between putative parents (Sun et al. 2002).

Speciation via hybridization, especially accompanied by polyploidization, plays role in the evolution of angiosperms. Hybridization can induce rapid genomic changes and thereby cause changes in DNA content, including gain or loss of DNA. Based on sequences of the internal transcribed spacer (ITS) of nuclear ribosomal DNA, the origin and evolution of the putative hybrid species *H. goniocarpa* were investigated by Sun et al. (2003). The ITS sequences of eight samples from the putative hybrid species and its possible and sympatrically distributed parental species *H. rhamnoides* ssp. *sinensis* and *H. neurocarpa* had 58 variable sites among these three species. The observed variable sites were all different between *H. rhamnoides* ssp. *sinensis* and *H. neurocarpa*. In the four samples of *H. goniocarpa*, 91.4% of the variable sites showed perfect nucleotide additivity of the parental species, and the remaining five sites showed concerted evolution to some degree. ITS sequencing results strongly supported that *H. goniocarpa* was derived through hybridization between *H. rhamnoides* ssp. *sinensis* and *H. neurocarpa*.

The findings of Sun et al. (2003) were further provided strength by the findings of Zhou et al. (2010). The nuclear DNA content of a diploid hybrid species, *Hippophae goniocarpa*, and its two parental species was examined. The nuclear 2C DNA value was 2.95 $\pm$ 0.08 pg for *H. goniocarpa*; for the parental species, it ranged from 2.54 $\pm$ 0.09 pg to 2.66 $\pm$ 0.12 pg for *H. rhamnoides* ssp. *sinensis* and from 3.15 $\pm$ 0.19 pg to 3.49 $\pm$ 0.06 pg for *H. neurocarpa* ssp. *neurocarpa*. The nuclear DNA content of *H. goniocarpa* was intermediate between those of the two parental species, confirming that this species is of diploid hybrid origin and further suggesting that it is in an early stage of the process of speciation.

To discriminate between medicinal and non-medicinal *Hippophae* species by DNA barcodes, the ITS2, *psbA-trnH*, and a combination of ITS2 and *psbA-trnH* (ITS2 + *psbA-trnH*) regions were used as candidate barcodes by Liu et al. (2015). DNA was extracted from the dried fruit samples. Primer pairs ITS2F/3R for ITS2 and *psbA*F/*trnH*R for psbA-trnH were used for PCR amplification. The purified PCR products were bidirectionally sequenced. Genetic distances were calculated according to the Kimura two-parameter model, and phylogenetic tree was constructed based on neighbor-joining (NJ) method; bar-coding gap was also analyzed to assess identification efficiency. Amplification and sequencing efficiencies for both ITS2 and *psbA-trnH* were 100%. Sequence data revealed that ITS2 + *psbA-trnH* were the most suitable candidate barcode at the species and subspecies level. The closely related *Hippophae* species were effectively differentiated in the NJ tree. It suggested that the combination of the two loci, ITS2 + *psbA-trnH*, is applicable to the identification of medicinal and non-medicinal *Hippophae* species (Liu et al. 2015).

Indian *Hippophae* species and subspecies from northwestern Himalayan region of India were authenticated, and phylogenetic relationship between *Hippophae* species and subspecies was established by using internal transcribed spacer (ITS) region markers (Jadhav and Sharma 2014). ITS marker authenticated ecotypes from Lahaul region as *H. rhamnoides* subspecies *turkestanica* with 98% identity while ecotypes from Spiti, Kinnaur, and Leh region as *H. salicifolia* with 97–98% identity. Phylogenetic studies on the ecotypes from four different regions of northwestern Himalayas revealed three major clades, namely, *H. rhamnoides* subspecies *turkestanica* and *H. salicifolia*, and followed both species specificity and geographic affiliation; an out-group clade of *Elaeagnus* spp. was also obtained. The molecular phylogenetic trees indicated that most of *Hippophae* species and subspecies are closely related and share common clade, while all the out-groups have separate clade. The observations clearly showed the efficiency of ITS markers as molecular tools for species authentication and establishing phylogenetic relationship in *Hippophae*.

9 Conclusion

The genetic biodiversity in high-altitude-growing sea buckthorn plant results in large variation in bioactive molecules responsible for medicinal properties. Hence, there is need for proper identification of plant species used for medicinal and scientific research purposes. Various methods have been used, including molecular markers and bar-coding, for sea buckthorn identification and plant sex determination. Plant characterization using DNA-based techniques will also be useful for efficient engineering of highly productive cultivars with wide range of adaptations to different environments of cultivation. The generated markers could also distinguish the sea buckthorn products from possible adulterants.

References

Agarwal M, Shrivastava N, Padh H (2008) Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Rep 27:617–631

Alam Z (2004) Important therapeutic uses of sea buckthorn (Hippophae): a review. J Biol Sci 4:687–693

Ambaye RY, Khanolkar VR, Panse TB (1962) Studies on tumor inhibitory activity of indigenous drugs. Part I. Tumor inhibitory activity of *Hippophae salicifolia* D. Don. Proc Ind Acad Sci 56 (Sec. B):123–129

Bartish IV, Jeppsson N, Bartish GI (2000) Inter- and intra-specific genetic variation in *Hippophae* (Elaeagnaceae) investigated by RAPD markers. Plant Syst Evol 225:85–101

CBOL Plant Working Group (2009) A DNA barcode for land plants. Proc Natl Acad Sci 106:12794–12797

Chawla A, Stobdan T, Srivastava RB et al (2015) Sex-biased temporal gene expression in male and female floral buds of seabuckthorn (*Hippophae rhamnoides*). PLoS One 10:e0124890

Chen S, Yao H, Han J et al (2010) Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS One 5:e8613

Chen S, Pang X, Song J et al (2014) A renaissance in herbal medicine identification: from morphology to DNA. Biotechnol Adv 32:1237–1244

Chowdhury MA, Andrahennandi CP, Slinkard AE, Vandenberg A (2001) RAPD and SCAR markers for resistance to ascochyta blight in lentil. Euphytica 118:331–337

Das K, Ganie SH, Mangla Y et al (2016) ISSR markers for gender identification and genetic diagnosis of Hippophae rhamnoides ssp. turkestanica growing at high altitudes in Ladakh region (Jammu and Kashmir). Protoplasma 254(2):1063–1077. https://doi.org/10.1007/s00709-016-1013-8

Ding J, Ruan CJ, Guan Y et al (2016) Characterization and identification of ISSR markers associated with oil content in sea buckthorn berries. Genet Mol Res 15(3). https://doi.org/10.4238/gmr.15038278

Divekar HM, Shyam R, Gupta AK, Pahwa ML, Grover SK, Kumar R, Srivastava KK (1996) Screening of adaptogenic activity of some Laddakhi plants. Report no. 1/96 DIPAS (Min. of Def.), Delhi, India

Durand R, Durand B (1990) Sexual determination and sexual differentiation. Crit Rev Plant Sci 9:295–316

Ercili S, Orhan E, Yildirim N, Agar G (2008) Comparison of seabuckthorn genotypes (*Hippophae rhamnoides* L.) based on RAPD and FAME data. Turk J Agric Forestry 32:363–368

Frank-Thorsten K (2004) Parataxonomy vs. taxonomy in biodiversity studies- pitfalls and applicability of 'morphospecies' sorting. Biodivers Conserv 13:795–812

Geetha S, Gupta A (2011) Medicinal and therapeutic potential of Seabuckthorn (*Hippophae rhamnoides* L.). J Ethnopharmacol 138:268–278

Geetha S, Sai Ram M, Singh V, Ilavazhagan G, Sawhney RC (2002) Anti-oxidant and immunomodulatory properties of sea-buckthorn (*Hippophae rhamnoides*) – an *in vitro* study. J Ethnopharmacol 79:373–378

He C, Gao G, Zhang J et al (2016) Proteome profiling reveals insights into cold-tolerant growth in sea buckthorn. Proteome Sci 14:14

Heath D, Williams RD (1980) Life at high altitude. Arnold- Heinemann Publishers (India) Pvt. Ltd., New Delhi

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proc R Soc Lond B 270:313–321

Hernández P, Martín A, Dorado G (1999) Development of SCARs by direct sequencing or RAPD products: a practical tool for the introgression and marker-assisted selection of wheat. Mol Breed 5:245–253

Hilu KW, Liang H (1997) The matK gene: sequence variation and application in plant systematics. Am J Bot 84:830–839

Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. PLoS One 6:e19254

Jaakola L, Suokas M, Häggman H (2010) Novel approaches based on DNA barcoding and high-resolution melting of amplicons for authenticity analyses of berry species. Food Chem 123:494–500

Jadhav MS, Sharma TR (2014) Molecular authentication of Hippophae species and subspecies and their phylogenetic analysis based on nuclear ITS sequences from north western Himalayan region of India. Res J Biotechnol 9:62–68

Kallio H, Yang B, Peippo P (2002a) Effects of different origins and harvesting time on vitamin C, tocopherols and tocotrienols in sea-buckthorn (*Hippophae rhamnoides* L.) berries. J Agric Food Chem 50:6136–6142

Kallio H, Yang B, Peippo P, Tahvonen R, Pan R (2002b) Triacylglycerols, glycerophospholipids, tocopherols, and tocotrienols in berries and seeds of two subspecies (ssp. sinensis and mongolica) of Sea Buckthorn (Hippophaë rhamnoides). J Agric Food Chem 50:3004–3009

Korekar G, Sharma RK, Kumar R et al (2012) Identification and validation of sex-linked SCAR markers in dioecious Hippophae rhamnoides L. (Elaeagnaceae). Biotechnol Lett 34:973–978

Korekar G, Dolkar P, Singh H et al (2013) Genotypic and morphometric effect on fruit oil content in seventeen natural population of seabuckthorn (*Hippophae rhamnoides* L.) from Trans-Himalaya. Natl Acad Sci Lett 36:603–617

Korekar G, Dolkar P, Singh H et al (2014) Variability and the genotypic effect on antioxidant activity, total phenolics, carotenoids and ascorbic acid content in seventeen natural population of Seabuckthorn (Hippophae rhamnoides L.) from trans-Himalaya. Food Sci Technol 55:157–162

Kress WJ, Wurdack KJ, Zimmer EA et al (2005) Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci U S A 102:8369–8374

Kumar R, Pasha MAQ, Khan AP, Gupta V, Grover SK, Norboo T, Srivastava KK, Selvamurthy W, Brahamchari SK (2003) Association of high altitude systemic hypertension with the deletion allele of the angiotensin converting enzyme (ACE) gene. Int J Biometeorol 48:10–14

Kumar R, Pasha MAQ, Khan AP, Gupta V (2004) Renin angiotensin aldosterone system and ACE I/D gene polymorphism in high altitude pulmonary edema. Aviat Space Environ Med 75:981–983

Lahaye R, van der Bank M, Bogarin D et al (2008) DNA barcoding the floras of biodiversity hotspots. Proc Natl Acad Sci U S A 105:2923–2928

Liu Y, Sun W, Liu C et al (2015) Identification of *Hippophae* species (Shaji) through DNA barcodes. Chin Med 10:28

Mathew D, Parimelazhagan T, Gomez S, Ahmed Z (2007) Characterization of Seabuckthorn (*Hippophae* spp.) genetic resources in India using morphological descriptors. PGR Newsl 149:22–26

Mishra RN (2011) Rasayan – the ayurvedic perspective. Res J Pharm, Biol Chem Sci 2:270–282

Mishra P, Kumar A, Nagireddy A et al (2016) DNA barcoding: an efficient tool to overcome authentication challenges in the herbal market. Plant Biotechnol J 14:8–21

Persson HA, Nybom H (1998) Genetic sex determination and RAPD marker segregation in the dioecious species sea buckthorn (Hippophae rhamnoides L.). Hereditas 129:45–51

Purushothaman J, Suryakumar G, Shukla D, Malhotra AS, Kasiganesan H, Kumar R, Sawhney RC, Chami A (2008) Modulatory effects of seabuckthorn (Hippophae rhamnoides L.) in hypobaric hypoxia induced cerebral vascular injury. Brain Res Bull 77:246–252

Purushothaman J, Suryakumar G, Shukla D, Jayamurthy H, Kasiganesan H, Kumar R, Sawhney RC (2011) Modulation of hypoxia-induced pulmonary vascular leakage in rats by seabuckthorn (Hippophae rhamnoides L.). Evid Based Complement Alternat Med 2011:574524

Ramachandran U, Divekar HM, Grover SK, Srivastava KK (1990) New experimental model for evaluation of adaptogenic products. J Ethanopharmacol 29:275–281

Reddy BU (2009) Molecular phylogeny of angio-spermic plant families using rbcL gene sequences. Int J Bioinf Res 1:27–36

Ruan CJ, Li DQ (2005) AFLP fingerprinting analysis of some cultivated varieties of seabuckthorn (*Hippophae rhamnoides*). J Genet 84:311–316

Ruan C, Qin P, Zheng J, Hea Z (2004) Genetic relationships among some cultivars of seabuckthorn from China, Russia and Mongolia based on RAPD analysis. Sci Hortic 101:417–426

Ruan CJ, Li H, Mopper S (2009) Characterization and identification of ISSR markers associated with resistance to dried-shrink disease in seabuckthorn. Mol Breed 24:255–268

Saggu S, Kumar R (2007a) Possible mechanism of adaptogenic activity of sea buckthorn (Hippophae rhamnoides) during exposure to cold, hypoxia and restraint (C-H-R) stress induced hypothermia and post stress recovery in rats. Food Chem Toxicol 45:2426–2433

Saggu S, Kumar R (2007b) Modulatory effect of sea buckthorn leaf extract on oxidative stress parameters in rats during exposure to cold, hypoxia and restraint (C-H-R) stress and post stress recovery. J Pharm Pharmacol 59:1739–1745

Saggu S, Kumar R (2008) Effect of sea buckthorn leaf extracts on circulating energy fuels, lipid peroxidation and antioxidant parameters in rats during exposure to cold, hypoxia and restraint (C-H-R) stress and post stress recovery. Phytomedicine 15:437–446

Saggu S, Gupta V, Sawhney RC, Rai PK, Kumar R (2006) Analysis of heavy metals in herbal extracts of high altitude growing plants Hippophae rhamnoides and Rhodiola imbricata. Toxicol Int 13:111–117

Saggu S, Divekar HM, Gupta V, Sawhney RC, Banerjee PK, Kumar R (2007) Adaptogenic and safety evaluation of sea-buckthorn (*Hippophae rhamnoides*) leaf extract: a dose dependent study. Food Chem Toxicol 45:609–617

Savelkoul PHM, Aarts HJM, de Haas J et al (1999) Amplified-fragment length polymorphism analysis: the state of an art. J Clin Microbiol 37:3083–3091

Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopaedia of life: an introduction to DNA bar-coding. Philos Trans R Soc Lond Ser B Biol Sci 360:1805–1811

Saxena S, Kumar R, Madan T, Gupta V, Murlidhar K, Sharma PU (2005) Association of polymorphisms in pulmonary surfactant protein A1 and A2 genes with high altitude pulmonary edema. Chest 128:1611–1619

Shah AH, Ahmad SD, Khaliq I et al (2009) Evaluation of phylogenetic relationship among sea buckthorn (Hippophae Rhamnoides L spp. Turkestanica) wild ecotypes from Pakistan using amplified fragment length polymorphism (AFLP). Pak J Bot 41:2419–2426

Sharma SK, Chunekar KC (1998) Medicinal plants used in Ayurveda. National Academy of Ayurveda, Rastriya Ayurveda Vidyapeeth Publication, New Delhi

Sharma N, Gupta S, Pathak R et al (2015) Genetic divergence among seabuckthorn (*Hippophae sps.* (L)) using molecular marker based analysis with respect to seed protein and anti-oxidant properties. Adv Anim Vet Sci 3:211–224

Silva HCD, Caraciolo RLF, Marangon LC et al (2014) Evaluating different methods used in ethnobotanical and ecological studies to record plant biodiversity. J Ethnobiol Ethnomed 10:48

Singh R, Mishra SN, Dwivedi SK, Ahmed Z (2006) Genetic variation in seabuckthorn (*Hippophae rhamnoides* L.) population of cold arid Ladakh (India) using RAPD markers. Curr Sci 91:1321–1322

Soni HB, Dabhi M, Thomas S (2013) Perspective on phytochemical and biochemical compounds of selected Indian medicinal plants. Int Res J Chem 1:37–45

Sougrakpam Y, Deswal R (2016) *Hippophae rhamnoides* N-glycoproteome analysis: a small step towards sea buckthorn proteome mining. Physiol Mol Biol Plants 22:473–484

Srihari JM, Verma B, Kumar N et al (2013) Analysis of molecular genetic diversity and population structure in sea buckthorn (*Hippophae spp* L.) from north-western Himalayan region of India. J Med Plant Res 7:3183–3196

Stobdan T, Chaurasia OP, Korekar G et al (2010) Attributes of seabuckthorn (Hippophae rhamnoides L.) to meet nutritional requirements in high altitude. Def Sci J 60:226–230

Stobdan T, Yadav A, Mishra GP, Chaurasia OP, Srivastava RB (2011) Seabuckthorn: the super plant (production, characterization, postharvest & health applications). Defence Institute of High Altitude Research, Defence Research and Development Organization, Leh-Ladakh, India

Stobdan T, Korekar G, Srivastava RB (2013) Nutritional attributes and health application of seabuckthorn (*Hippophae rhamnoides* L.) – a review. Curr Nutr Food Sci 9:151–165

Sun K, Chen X, Ma R et al (2002) Molecular phylogenetics of *Hippophae* L. (Elaeagnaceae) based on the internal transcribed spacer (ITS) sequences of nrDNA. Plant Syst Evol 235:121–134

Sun K, Ma RJ, Chen XL (2003) Hybrid origin of the diploid species *Hippophae goniocarpa* evidenced by the internal transcribed spacers (ITS) of nuclear rDNA. Belg J Bot 136:91–96

Sun K, Chen W, Ma R et al (2006) Genetic variation in *Hippophae rhamnoides* ssp. *sinensis* (Elaeagnaceae) revealed by RAPD markers. Biochem Genet 44:186–197

Tang X (2002) Intrinsic change of physical and chemical properties of sea buckthorn (Hippophae rhamnoides) and implications for berry maturity and quality. J Hortic Sci Biotechnol 77:177–185

Tang X, Tigerstedt PMA (2001) Inheritance of flowering, maturity, fruit yield, and winter hardiness of sea buckthorn (Hippophae rhamnoides). J Am Soc Hort Sci 126:744–749

Techen N, Parveen I, Pan Z, Khan IA (2014) DNA barcoding of medicinal plant material for identification. Curr Opin Biotechnol 25:103–110

Tulsawani R, Sharma P, Divekar HM, Meena RN, Singh M, Kumar R (2010) Supplementation of fruit extract of *Hippophae rhamnoides* speeds adaptation to simulated high altitude stressors in rats. J Complement Integr Med 7:Article 30

Varshney RK, Grane A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55

Vassou SL, Nithaniyal S, Raju B, Parani M (2016) Creation of reference DNA barcode library and authentication of medicinal plant raw drugs used in Ayurvedic medicine. BMC Complement Altern Med 16(Suppl 1):186

Yang B, Kallio HP (2001) Fatty acid composition of lipids in sea buckthorn (*Hippophaë rhamnoides* L.) berries of different origins. J Agric Food Chem 49:1939–1947

Yao Y, Tigerstedt PMA (1993) Isozyme studies of genetic diversity and evolution in *Hippophae*. Genet Resour Crop Evol 40:153–164

Yao Y, Tigerstedt PMA (1994) Genetic diversity in Hippophae L. and its use in plant breeding. Euphytica 77:165–169

Yao Y, Tigerstedt PMA (1995) Geographical variation of growth rhythm, height, and hardiness, and their relations in Hippophae rhamnoides. Am Soc Hort Sci 120:691–698

Zhou X, Ma J, Wang W et al (2010) Genome size of the diploid hybrid species Hippophae goniocarpa and its parental species H. rhamnoides ssp. Sinensis and H. neurocarpa spp. Neurocarpa (Elaeagnaceae). Acta Biol Cracov Ser Bot 52(2):12–16

Closing Shots: DNA Barcoding and Molecular Phylogeny

Subrata Trivedi, Hasibur Rehman, Shalini Saggu, Chellasamy Panneerselvam, and Sankar K. Ghosh

Abstract Conservation and assessment of bio-resources is a challenging task. With the rapid advancement of applied molecular biology techniques and bioinformatics, DNA barcoding and molecular phylogeny studies have gained momentum in the past decade. This chapter gives a brief description of all the chapters in this book—DNA Barcoding and Molecular Phylogeny. It includes some important case studies along with DNA barcoding information in relation to invasive alien species, forensic botany, infection management, plant animal interactions, fishery management, DNA database management, etc. Other topics include DNA barcoding and molecular phylogeny in microbes, aquatic plants, medicinal plants, red alga, ichthyoplankton, rays, fishes, reptiles, birds, and ruminant mammals. The concept of R as a bioinformatics tool in DNA barcoding is also discussed.

Keywords DNA barcoding · Molecular phylogeny · Biodiversity · Invasive alien species · Ruminant mammals · Genetic variation · Bioinformatics

DNA barcoding has become a major focus of research in recent time and has gained global attention. In this book, we have covered broad aspects of DNA barcoding and

S. Trivedi (✉) · C. Panneerselvam
Department of Biology, University of Tabuk, Tabuk, Saudi Arabia
e-mail: strivedi@ut.edu.sa

H. Rehman
Department of Biology, University of Tabuk, Tabuk, Saudi Arabia

Departments of Pathology, School of Medicine, The University of Alabama at Birmingham (UAB), Birmingham, AL, USA

S. Saggu
Department of Biology, University of Tabuk, Tabuk, Saudi Arabia

Departments of Dermatology, School of Medicine, The University of Alabama at Birmingham (UAB), Birmingham, AL, USA

S. K. Ghosh
University of Kalyani, Nadia, West Bengal, India

© Springer International Publishing AG, part of Springer Nature 2018

S. Trivedi et al. (eds.), *DNA Barcoding and Molecular Phylogeny*,
https://doi.org/10.1007/978-3-319-90680-5_27

molecular phylogeny along with some case studies. The first chapter highlights the significance and utilities of this comparatively new concept. This chapter reveals the different gene markers that are used for DNA barcoding in various groups of living organisms. In the post-genomics era, this technique has several practical applications. Metabarcoding and NGS technology are also discussed. This chapter also highlights how DNA barcoding can be used as a quality control tool.

The second chapter deals with R in DNA barcoding. This is a new topic, and very little information is available in this arena at present. This is a significant development in the assessment of DNA sequences. This bioinformatics tool has been discussed in this chapter. We especially invited this chapter on R language to enrich our book on DNA barcoding and molecular phylogeny.

The third chapter deals with the implications of DNA barcoding. DNA barcoding has become a promising tool for validating the different species and also has several implications like detection of mislabeling in the food industry, safety assessment, controlling agricultural pests, detection of disease vectors, monitoring water quality, etc.

The fourth chapter deals with the prospects and limitations of DNA barcoding in birds. Museums contain many extinct birds both as whole specimens and also as skin and feathers. DNA from such specimens is used to identify avian species along with presently available species.

Invasive alien species (IAS) is a topic of great interest not only for ecological consequences but also for the massive economical damage caused by the introduction of these species. The fifth chapter of this book is devoted to this vital topic on invasive species identification through DNA barcoding. This chapter not only highlights how DNA barcoding can promptly identify the invasive species but also in the management process of invasive species.

It is well known that microbes are available in every corner of our planet and even in extreme conditions. The sixth chapter in this book elucidates the prospects of discovery and identification of microbes through DNA barcoding.

The next chapter (Chap. 7) deals with a topic in medical science—how DNA barcoding of bacteria can help in infection management. This chapter deals with the application of DNA barcoding in clinical microbiology.

Various ecological processes are involved in plant-animal interactions. Any disruption in this interaction can be detrimental not only for the species involved but also other species in the community. Traditionally these interactions have been studied by field observations. Chapter 8 of this book demonstrates how DNA barcoding has revolutionized the field of community ecology.

The ninth chapter of this book is about the nexus between DNA barcoding and forensic botany. Forensic botany can provide supporting evidence during court investigations. DNA barcoding can provide accurate identification platform for solving such cases.

Red alga has immense ecological and economical value. Chapter 10 is devoted to the molecular identification of red alga through DNA barcoding.

Large numbers of DNA sequences have been continuously deposited to global databases. Chapter 11 deals with the different DNA databases of the world. Promises and limitations of these databases are discussed in relevance to DNA barcoding.

Assessment of plant diversity is a challenging task, particularly in the aquatic and marine ecosystems. Chapter 12 of this book explains how DNA barcoding can serve as a useful tool for the assessment of aquatic plant biodiversity.

Mosquitoes act as major vectors for several epidemics around the world like malaria, dengue, Zika viral disease, etc. Mosquito Barcode Initiative is a major global barcoding project devoted to barcoding of mosquitoes worldwide. Chapter 13 relates to the DNA barcoding of insects with special emphasis on the barcoding of mosquitoes.

Identification of rays is a challenging task due to overlapping morphological characters. Chapter 14 relates to DNA barcoding of rays from South China Sea. This chapter additionally addressed the management plan of elasmobranch fishery in Malaysia.

The next chapter is on the complete molecular phylogeny of elasmobranchs with emphasis on the phylogeny for framing conservation strategies. Fish is an important protein course, and fishery industry provides huge employment worldwide. Chapter 16 is a review of the DNA barcoding and fish taxonomy in India. Chapter 17 of this book is on the applications of DNA barcoding in fisheries. Although reptiles show significant diversity, only a few DNA barcoding studies have been conducted on reptiles as compared to other vertebrate groups. The 18th chapter discusses the DNA barcoding scenario in different orders of class Reptilia. The next chapter is regarding the taxonomically wide biogeographic study of birds in the context of DNA barcoding.

Ruminant mammals differ from other mammals by having four-chambered stomach. Chapter 20 is on the molecular characterization of ruminant mammals. This chapter depicts how the *COI* gene can be used for the molecular identification of ruminants.

Marine actinobacterium is a huge source for bioactive compounds including secondary metabolites. Chapter 21 involves study on 16s rDNA sequences for the molecular identification and phylogeny studies for culturable marine actinobacteria collected from Andamans in India.

Chapter 22 is an important case study in which DNA barcoding and molecular phylogeny are done on the bacterial species isolated from different tissues from three commercial tropical tidal fishes of Malaysia.

The next chapter is on DNA barcoding and molecular phylogeny of ichthyoplankton and juvenile fishes of Kuantan River in Malaysia. This study may be valuable for analyzing post-flood effect on fish distribution in tropical river and implementing plans for future fishery resource management.

Genetic variation is essential for long-term viability of a population because it increases the adaptability of the species with the changing environment. In the hatcheries, loss of genetic variation due to inbreeding is a serious concern. Chapter 24 elucidates this issue in the context of DNA barcoding. Here the genetic variation in both wild and hatchery populations of *Anabas testudineus* is revealed

which can be very useful for future sustainable selective breeding programs. The deserts of Arabian Peninsula are ideal habitat for several species of reptiles. Chapter 25 is a case study on the reptiles of Tabuk province of Saudi Arabia. Chapter 26 is on molecular characterization and DNA barcoding of a high altitude medicinal plant, *Hippophae rhamnoides.*

The concept of DNA barcoding and molecular phylogeny has gained tremendous attention, and several researchers and academicians around the globe are actively involved in this current and important arena of science. At present, only a very few books are available on DNA barcoding. This book will provide a diverse account of selected research projects (DNA barcoding, molecular phylogeny, biodiversity, evolution, and molecular taxonomy) and case histories from sites and/or laboratories around the globe working on DNA barcoding and molecular phylogeny.